THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Ecological Speciation in *Senecio lautus*

Maria Clara Melo Hurtado

MSc

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2014

School of Biological Sciences

Abstract

The role of natural selection in driving speciation between interbreeding populations has been one of the most controversial topics in evolutionary biology. Populations that inhabit contrasting environments can evolve adaptive traits in the local habitat that can reproductively isolate them. This process seems straight forward between allopatric populations, where populations isolated by geographic barriers can accumulate adaptive genetic differentiation and evolve reproductive isolation, but when populations are in the same or proximate localities, gene flow may oppose divergence and speciation. Although speciation with gene flow is an increasingly supported phenomenon, studies still face complications dissecting the effect of divergence time vs. gene flow, even with extensive molecular data. An increasing number of examples from nature provide evidence for this model of speciation. However, most of the studies have been conducted in animal systems, leaving ecological speciation in plants largely unexplored. Dune and Headland populations within the *Senecio lautus* ecotype and species complex occur proximate to each other in several coastal localities of Australia, displaying very contrasting morphologies despite being interfertile. Supported by a robust phylogenetic study, each Dune and Headland pair shows an independent origin and it displays characteristics that suggest an important role for ecology in the diversification of its ecotypes. Here, I used a combination of ecological, molecular and comparative approaches to investigate the process of speciation in the Australian groundsel *Senecio lautus.* In reciprocal transplants in the field and experiments in the glasshouse I found that Dune and Headland populations are strongly isolated by ecology based reproductive barriers and that intrinsic barriers contribute little to it. Then I discovered that most parapatric pairs displayed drastic reductions in gene flow, while more distant populations from the same ecotype still exchange genes. Finally, I provide evidence that the multiple Dune and Headland pairs in the system are evolving under the model of ecological parallel speciation thus providing strong evidence for the role of natural selection in plant speciation. The experimental results in my dissertation suggest that ecology is not only able to counteract gene flow at early stages of divergence, but that can also take populations to the most advanced stages of speciation where populations no longer exchange genes in the field. These results strengthen previous studies that suggest that the evolution of intrinsic reproductive isolation may be decoupled from the process of speciation. *Senecio lautus* constitutes the first well-supported case for the parallel ecological speciation in plants, and provides an excellent opportunity to study speciation with gene flow.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the *Copyright Act 1968*.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Peer-reviewed papers

Melo, M. C., Grealy, A., Brittain, B., Walter, G. M. and Ortiz-Barrientos, D. (2014), Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsel. New Phytologist, doi: 10.1111/nph.12779.

Publications included in this thesis

Melo, M. C., Grealy, A., Brittain, B., Walter, G. M. and Ortiz-Barrientos, D. (2014), Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsel. New Phytologist, doi: 10.1111/nph.12779 – as Chapter II.

Contributor	Statement of contribution
Maria C. Melo	Reciprocal transplant experiments design 60% Data collection 70% Data analyses 80% Writing 70%
Alicia Grealey	Reciprocal transplant experiments design 20% Data collection 20%
Beth Brittain	Transplant experiment conduction 20% Data collection 10%
Greg Walter	Sample collection 100%
Daniel Ortiz-Barrientos	Reciprocal transplant experiments design 20% Data analyses 20% Writing 30%

Contributions by others to the thesis

Contributor	Specific contribution
Maddie E. James	Contributed to the design, execution and analysis of Access array experiment.
Melanie J. Wilkinson	Contributed to the design, execution and analysis of Access array experiment.
Diana M. Bernal	Contributed to the design, execution and analysis of Access array experiment.
Federico Roda	Contributed to the design, execution and analysis of Access array experiment.
Thomas Richards	Creation of population crosses for H5 and D4
Daniel Ortiz-Barrientos	Contributed to Experimental Design, Sample collection, Data analysis, and Writing

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Acknowledgements

I would like to express my special appreciation and thanks to my supervisor Daniel Ortiz-Barrientos. Daniel thanks for offering me the best guidance in all of the aspects that involve carrying out high quality scientific research. Thanks for your invaluable friendship and support, and for making of the lab a very friendly and happy environment. I would also like to thank my committee members. L. Cook and S. Chenoweth gave me important advise in all the milestones of my thesis. I also want to thank S. Blomberg for important statistical advice.

Special thanks to A. Posada for all her help and generosity; she made my arrival to Autralia very smooth and homey. J. Donohoe for his help, friendship and making countable the uncountable. I would like to thank everyone in my lab for sharing very important and special moments. Each one of you provided me with help and support whenever I needed. I want to specially thank Huanle Liu for his constant and peaceful company along these years. I also want to thank all the people that work in the School of Biological sciences, who in many ways made this process a very enjoyable experience. I want to thank all of my friends here in Australia, I have found in all off invaluable support. Finally I want to thank all my Family, I love you so much.

Keywords

Speciation, paratry, local adaptation, parallel evolution, extrinsic reproductive isolation, intrinsic reproductive isolation

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060311, Speciation and Extinction, 40%

ANZSRC code: 060411, Population, Ecological and Evolutionary Genetics, 30%

ANZSRC code: 060303, Biological Adaptation, 30%

Fields of Research (FoR) Classification

FoR code: 0602, Ecology, 30%

FoR code: 0603, Evolutionary Biology, 40%

FoR code: 0604, Genetics, 30%

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2N0m0>1	Population migration rate from population 2 to population 1
2N1m1>0	Population migration rate from population 1 to population 2
AMOVA	Analisis of molecular variance
Anova	Analisis of variance
CPP	Conspecific pollen precedence
D	Dune
DH	Divergence hitchhiking
DS	Direct selection
F1-D	F1 hybrid with Dune mother
F1-H	F1 hybrid with Headland mother
FF	Fixed-fixed
FP	Fixed-polymorphic
GH	Genome hitchhicking
H	Headland
IBA	Isolation by adaptation
IBD	Isolation by distance
m	Migration rate
m0>1	Migration rate from population 2 to populaiton 1
m1>0	Migration rate from population 1 to population 2
N0	Effective population sizes of population 1
N1	Effective population sizes of population 2
N2	Effective population sizes of ancestral population
Ne	Effective population size
PF	Polymorphic-fixed
PP	Polymorphic-polymorphic
q0	Population size of population 1
q1	Population size of population 2
q2	Population size of hypothetical ancetral population
QTL	Quantitative trait loci
RI	Reproductive isolation
s	Selection
S.E.	Standard error
to	Divergence time

CHAPTER I

GENERAL INTRODUCTION

Ecological speciation or the evolution of reproductive isolation (RI) between populations adapting to contrasting environments is still a poorly understood process in evolutionary biology, particularly when it occurs in the face of gene flow (Nosil, 2012). Although Darwin's model of speciation was sympatric and driven by competition, the founders of the new synthesis believed that ecology could be important, but mostly between populations diverging in allopatry (Dobzhansky, 1940; reinforcement being the great exception; Mayr, 1963; Coyne & Orr, 2004). Only recently the role of ecology on speciation has been settled, but the road was long and sinuous starting with Darwin, and currently culminating with the conceptual demarcation of adaptive radiation and ecological speciation (Schluter, 2001; Rundle & Nosil, 2005).

While Darwin showed that natural selection was a sufficient process to explain diversity on earth, he did not elaborate on how adaptive forms differed from species as biodiversity units. For Darwin, varieties, races, and species were to great extent exchangeable terms. His view was, and remains adequate to match the adaptive process with our intuitive way of classifying organisms based on morphological differences. However, trait evolution by natural selection could not explain why reproduction failed between certain species crosses. It was unconceivable that natural selection could favor the evolution of traits that led to sterility or inviability. But this was perhaps much to ask from Darwin, as his knowledge of genetics was erroneous at best.

After Mendel's laws were rediscovered, and the founders of the new synthesis expanded this simple mathematics to the population level, it became also clear that these concepts could be applied to our understanding of speciation. Theodosius Dobzhansky championed the idea that populations adapting to their environments would evolve mechanisms that would protect coadapted gene complexes from dissolving in the face of gene flow. He then coined the term reproductive isolating mechanisms. Note that technically the only mechanism of speciation that exists is that of reinforcement, or the evolution of enhanced prezygotic isolation in response to maladapted hybridization. In general, today we accept that reproductive barriers evolve as a byproduct of divergence, be adaptation or random via genetic drift.

Together with Dobzhansky, Ernest Mayr was instrumental in bringing to life the Biological Species concept and thus revolutionizing the study and understanding of the origin of new species. The reason is two fold: first, the study of speciation became synonymous with the evolution of RI, thus freeing (for better or worse) the process from the action of natural selection. And second, it solved Darwin's conundrum as it was able to link adaptation with speciation. Nevertheless, this link remained conceptual for decades because the evolution of RI did not require necessarily the action of natural selection, and also because students of speciation largely devoted themselves to understanding the origin of RI under laboratory conditions, thus disassociating the origin of the species from the interaction between genes and environment (but see the work in *Rhagoletis* and in Darwin's finches for remarkable exceptions during the second half of the XXth century).

Largely due to the influence of Mayr, the study of speciation focused on allopatric speciation for many decades. This idea was also fueled by theoretical work showing that recombination was the antagonist of natural selection (e.g., Felsenstein, 1981). However, recent theoretical and empirical work suggests that allopatric and non-allopatric (e.g., parapatric and sympatric) forms of speciation exist in nature and could be common. But possibly, Mayr was not off the mark completely: speciation with gene flow seems to happen because portions, or most of the genome, fail to experience gene flow. Thus, the crux of studying speciation in parapatry or sympatry is to find the mechanisms that reduce recombination between species and as a consequence facilitate divergence by genetic drift, as it happens in allopatry. Whether the evolution by drift that ensues is responsible for the evolution of further RI remains one of the most challenging problems in current speciation studies.

But not everything that is interesting is about how to facilitate divergence by genetic drift during speciation with gene flow. To fully understand the process we need to address some fundamental questions about both the ecology and genetics of the evolution of RI. For instance, we remain largely ignorant about the selective agents driving adaptation and the evolution of extrinsic RI. Similarly, although we know how selection reduces gene flow, we lack estimates of the strength of selection in the field, and thus theoretical models continue to live in a vacuum of empirical data. More problematic, if surprising, there are competing models for whether intrinsic or extrinsic RI start the speciation process, and how it is completed. Finally, although we have a general framework for understanding speciation with gene flow and ecological speciation, we still do not have enough data to compare divergence processes between plants and animals, thus questioning whether the very essence of our models is correct and general.

Here I provide a brief theoretical background on the basic concepts to understand the role of natural selection in the evolution of RI and speciation. They will provide some of the foundations that complement the introductions to each of my chapters. For instance, in Chapter II I experimentally explore the role of natural selection in creating extrinsic RI between parapatric populations in the field. Then, in Chapter III I estimate the amount of gene flow that characterizes parapatric divergence, and then I finish in Chapter IV with a direct test of the parallel ecological speciation hypothesis. My results suggest a new biological model for the study of ecological speciation with gene flow.

Forms of RI

RI barriers can be several and of different nature (Coyne & Orr, 2004). Premating isolating barriers, also known as prezygotic isolation are those that prevent individuals from different populations to mate and produce a fertilized zygote. Because they prevent gene flow early in the life cycle of individuals they have also been referred as early acting barriers. These barriers are mainly due to ecological factors and/or to sexual selection, both leading to assortative mating between individuals of the same population (Coyne & Orr, 2004). For instance in inland and coastal populations of the monkey flower *Mimulus guttatus,* individuals migrating to the alternative environment tend to find their physiology compromised to be able to survive and/or successfully mate. This reproductive barrier known as immigrant inviability (Nosil *et al.*, 2005), causes assortative mating amongst individuals adapting to the same environment in monkeyflowers (Lowry *et al.*, 2008b), and other systems such as aphids, walking sticks and fish (Via *et al.*, 2000; Nosil, 2004; Leinonen *et al.*, 2011). In *Heliconius* butterflies, wing colour pattern in addition of emitting warning of toxicity to predators, also serves as a cue for mating creating interbreeding amongst individuals that carry the same colour pattern (Jiggins *et al.*, 2001). Other barriers in this group such as mechanical isolation (e.g. genitals are not compatible between male and females of two diverging populations), mating system isolation (e.g. plants that evolve with self-fertilization and become isolated from their self-incompatible progenitor), and flowering time differences, also prevent interbreeding between individuals from different populations (Coyne & Orr, 2004).

In contrast, postmating isolating barriers are late acting barriers preventing gene flow after mating occurs. They can be postmating-prezygotic barriers, preventing fertilization after mating has occurred (e.g. gametic isolation due to pollen competition for ovules), and postzygotic barriers when fertilization is successful but gene flow is prevented in the hybrids (as in hybrid sterility or inviability). For instance, many species of plants show conspecific pollen precedence, or the ability

of conspecific pollen to outcompete heterospecific pollen grains when they both arrive to the same flower (Rieseberg *et al.*, 1995). In many studied insects, crosses between two species leads to strong reductions in hybrid fitness (Sturtevant, 1920; Lachaise *et al.*, 1986; Presgraves, 2002). Both prezygotic and postzygotic barriers are not exclusive to each other; they have actually been found to act jointly to reducing gene flow (Coyne & Orr, 2004).

We can also classify reproductive isolating barriers into intrinsic and extrinsic barriers, giving a sense of the evolutionary forces that may be causing them. When barriers to gene flow are intrinsic, genetic differences between species directly cause assortative mating, inviability or infertility in the hybrids. These phenotypes are independent of the environment in which the species encounter each other, and therefore are likely to arise from other evolutionary forces like genetic drift (Coyne, 1992). In contrast, when barriers to gene flow are extrinsic, factors from the environment cause the hybrids to be inviable or sterile, or they lead directly to assortative mating (Rundle & Whitlock, 2001; Rundle, 2002; Coyne & Orr, 2004). Although both intrinsic and extrinsic barriers can arise during ecological speciation (Agrawal *et al.*, 2011; Nosil, 2012), barriers that arise from the interaction of individuals with their external environment are more likely to evolve early in this process (ecology-based reproductive barriers), as they can be directly favoured by divergent natural selection (Schluter, 2001). Because these barriers are directly dependent on the environments, its been suggested that unlike intrinsic barriers, these barriers could disappear in cases of habitat disturbance as this would trigger species fusion. However, intrinsic barriers are perhaps as likely to disappear under secondary contact unless reinforcement evolves, a hybrid zone is established, or recombination-suppression mechanisms are already in place. It is surprising that the risk of species fusion has not been properly quantified under the existence of either extrinsic or intrinsic RI, and that it is largely assumed that reversibility is just related to ecology-based reproductive barriers.

Absolute and relative importance of RI

For many years the study of speciation focused on reproductive isolating barriers independently from what their contribution to speciation was (Schemske, 2010). An example of this is that intrinsic postzygotic barriers like hybrid sterility became classic barriers in the study of speciation, consequence of the ease to study postzygotic isolation under laboratory conditions in insects (Coyne & Orr, 1997). However, the importance of these barriers in speciation studies does not necessarily reflect their role and contribution to the evolution of total RI. Two major scenarios have been entertained before. First, intrinsic barriers can be very strong in their absolute effects, but they might relatively contribute very little to total RI when prezygotic barriers already do the bulk of the

job (see details below). Second, intrinsic barriers may have evolved long after the speciation process was completed. Therefore, linking the importance of postzygotic RI to the process of speciation goes beyond studying its presence, its genetics, and the genes underlying hybrid dysfunction. Despite this worry, it is very likely that a large fraction of intrinsic RI evolves before the completion of speciation. Contrastingly, the study of barriers product of the interaction of individuals with the environment have been either poorly studied, or heavily studied but in the context of adaptation and not speciation. For instance, reciprocal transplants in plants abound, but most of these studies were not framed in light of the evolution of extrinsic RI. This is surprising given the rich history of research on ecotype formation in plants, and the fact that extrinsic RI could have large effects on speciation. The clear-cut definition of ecological speciation, the emergence of the genotypic cluster species concept, and the ability to do genetics in the field are possibly some of the most important triggers of renewed interest in the role of extrinsic RI in speciation.

Because RI acts sequentially to reduce gene flow, reproductive isolating barriers that are early acting contribute more to the total isolation of diverging populations, than late acting ones (Ramsey *et al.*, 2003; Lowry *et al.*, 2008a). Coyne and Orr (1997) introduced this vision when they estimated the relative contribution of premating and intrinsic postzygotic isolation in *Drosophila.* Later, Ramsey and colleagues (2003) expanded this method to multiple reproductive barriers between *Mimulus lewisii* and *M. cardinalis,* finding that ecogeographic, pollinator, gametic and intrinsic postzygotic barriers were individually strong barriers, but when the relative contributions of each barriers were estimated, prezygotic barriers were stronger than postzygotic (Ramsey *et al.*, 2003). Lowr0y et al. (2008) reviewed the individual strength of different reproductive barriers on 19 cases of diverging populations/species in flowering plants. Their results suggested that although most reproductive barriers can prevent interbreeding between plant species, prezygotic reproductive barriers contributed the most to total reproductive divergence (Lowry *et al.*, 2008a). Similarly, Schemske (2010) found that prezygotic barriers contributed more to total RI than postzygotic barriers. This and other cases that found similar patterns drove the attention toward the power of natural selection in creating isolation between populations and the possible scenarios in which it can happen (Gavrilets, 2003; Mallet *et al.*, 2009; Sobel *et al.*, 2010)

Modes of speciation

Natural selection can drive speciation under two different geographic models: ecological speciation between geographically isolated populations (allopatry), and ecological speciation with gene flow (sympatry and parapatry, Gavrilets, 2003; Coyne & Orr, 2004; Mallet *et al.*, 2009; Sobel *et al.*,

2010; Nosil, 2012). If diverging populations are geographically isolated (allopatry), individuals that adapt to different environments can easily accumulate adaptive variation that favours the evolution of RI (Mayr, 1963; Felsenstein, 1981). Under this scenario genetic divergence occur across the entire genome of populations, and with genetic drift furthering differentiation. This is the simplest form of ecological speciation and perhaps the most common in nature (Nosil, 2012). The second model of ecological speciation involves populations that are adapting to contrasting environments and that are still exchanging genes. This model has been thought to be difficult as the homogenising effect of gene flow may constrain adaptive divergence and the evolution of RI. Although gene flow can theoretically impede ecological speciation (e.g. Felsenstein, 1981), over the past two decades there has been growing interest in the conditions facilitate it. Theoretical models demonstrate that under certain conditions (e.g. genetic linkage or chromosomal rearrangements), selection can overcome gene flow and result in population divergence (Gavrilets, 2004; Bolnick & Fitzpatrick, 2007; Thibert- Plante & Hendry, 2009). This and a recent growing list of strong empirical cases suggest that ecological speciation in the face of gene flow can be common (Fuller *et al.*, 2007; Sobel *et al.*, 2010). However, there is still an ongoing debate on the feasibility of this model mainly due to difficulty in detecting gene flow and distinguishing natural selection from other forces driving the speciation process (Nosil, 2012).

Gene flow during speciation

Detecting cases of speciation with gene flow is one of the most challenging goals in speciation (Coyne & Orr, 2004; Nosil, 2008). A first step is detecting the origin of the shared genetic variation between populations. For example in the cases of weak genetic differentiation it is difficult to distinguish between the homogenizing role of gene flow, and variation shared from the ancestor (when populations diverge, genetic variation can be shared for a long time, specially if populations are large and genetic drift is low, Muir & Schloetterer, 2005). If gene flow is the most likely explanation, it is also necessary to know the time in which it occurred. Gene flow after secondary contact (after an initial period of divergence in allopatry) creates similar patterns of genetic variation to speciation with gene flow, although divergent natural selection may have not driven speciation. Approaches to solve this problem attempt to estimate gene flow directly, compare levels of shared polymorphism between allopatric and parapatric populations, and to evaluate patterns of heterogeneous genomic divergence along the speciation continuum while comparing different scales of geographic isolation (Feder *et al.*, 2012; Feder *et al.*, 2013).

A combination of genes showing little divergence with others displaying strong genetic differentiation suggests a history of divergence with gene flow (Hey, 2006). This could be reflected in phylogenetic discordances amongst genes (Machado & Hey, 2003), significant levels of population migration rates in coalescent-based analysis (Isolation with Migration model, IM, Hey & Nielsen, 2004), and in contrasting patterns of heterogeneous genomic divergence between allopatric and parapatric populations (Nosil *et al.*, 2008; Nosil *et al.*, 2009a). For instance, in cave salamanders species with overlapping ranges, phylogenetic discordances and significant levels of gene flow estimated through IM analysis indicated that speciation in this group has faced the homogenising effects of gene flow (Niemiller *et al.*, 2008). Heterogeneous patterns of divergence driven by low recombining regions of the genome of M and S forms of *Anopheles gambiae,* are consistent with theoretical models that predict that differentiated regions harbouring important adaptive genes would remain differentiated while the rest of the genome keeps exchanging genes (Turner *et al.*, 2005). This and other examples are good candidates of speciation with gene flow.

Cases of parallel speciation

Recent studies on ecological speciation (Hatfield & Schluter, 1999; Foster *et al.*, 2007; Butlin *et al.*, 2008; Nosil *et al.*, 2008) usually compare multiple replicate cases of ecological divergence. This is perhaps the strongest approach to argue for the role of natural selection in speciation. In particular, systems with populations that present similar morphologies in similar environments are good systems where to study the evolution of RI driven by natural selection, or parallel ecological speciation. If populations are evolving in response to the similar selective pressures, RI is expected to evolve between populations adapting to contrasting environments. In contrast, it is expected that RI will not evolve between populations adapting to similar environments (mutation-order speciation would be an exception, Schluter, 2009; Ostevik *et al.*, 2012). Systems where this pattern occurs, besides showing strong evidence for ecological speciation, provide a unique opportunity to study diverging pairs at different stages of a continuous process of speciation (Fig. 1.1, Hendry, 2009). This benefits from viewing ecotypes/races/subspecies as evolutionary entities in a transition to become species, favouring a better understanding on how RI accumulates at different stages of the process (Clausen, 1951; Nosil *et al.*, 2009b). This view of speciation as a continuum does not imply that all ecotypes should become species or that the process is irreversible. For example a study that included eight stream-lake population transitions in sticklebacks, found variable progress towards ecological speciation ranging from weakly morphologically populations to very differentiated ecotypes (maybe already species), associated with different degrees of gene flow (Berner *et al.*, 2009). However, it is surprising that a limited number of cases of parallel ecological speciation are

found in animals, and almost absent in plants (candidate systems present weak evidence, Ostevik *et al.*, 2012).

The speciation continuum

The speciation continuum is a powerful idea where to examine how RI evolves in the face of gene flow. On one hand, speciation can start driven by local adaptation, and via the evolution of strong extrinsic RI. As gene flow comes to a halt in the field, genomes are free to evolve further differences that would normally be impossible to evolve in the face of gene flow. For instance, in this model intrinsic RI evolves later on during the speciation process once ecology based RI has reduced gene flow to levels that would not homogenise the populations. On the other hand, intrinsic RI can evolve in allopatry, and perhaps via a combination of genetic drift and natural selection. Once RI is very strong, populations can co-exist in sympatry and be sorted through ecological mechanisms, including ecological character displacement, or completing the speciation process through reinforcement of prezygotic isolation. Theoretical models for reinforcement are varied, and all suggest that reinforcement can evolve when recombination suppression mechanisms are in place (e.g., chromosomal inversions, Butlin, 2005), or the same allele spreads through the two populations and leads to assortative mating in each group (i.e., one-allele models, Felsenstein, 1981). Theories for the evolution of intrinsic RI following local adaptation are simple: local adaptation reduces gene flow locally in the genome, and through its effects on linked genes spreads genetic differentiation through genetic hitchhiking. Once these foci of differentiation are in place, further genetic differentiation can accumulate, including via genetic drift. Ultimately, it is expected that intrinsic RI could evolve. However, it is also possible that many loci of moderate effect govern subsequent adaptation, and thus it is theoretically possible that not only some small regions of the genome congeal, but also the entire genome via a process called genome hitchhiking (Feder *et al.*, 2012). The verdict on these models is still inconclusive, mostly because the study systems where genetics, ecology, and molecular biology can be performed across the speciation continuum remain scarce.

Speciation in the Senecio lautus *complex*

Senecio lautus is rapidly becoming an ideal model to study ecological speciation with gene flow. This Australian herb consists of multiple populations adapted to different environments, showing correlated morphologies with ecological conditions (Ali, 1966; Ali, 1968; Ali, 1969; Radford *et al.*,

2004; Thompson, 2005). Populations inhabiting the tablelands, woodlands, alpine meadows, sand dunes and rocky headlands (amongst others), exhibit contrasting morphologies that at first glance could suggest distant lineages. Instead they form a complex of closely related ecotypes that display high fertility under controlled conditions (Ali, 1964; Ornduff, 1964; Ali, 1966). The coastal system within *S. lautus* complex is particularly interesting: it consists of parapatric populations that exhibit contrasting morphologies found along the Australian coast. This parallel evolution of morphologies has been considered a signature of natural selection shaping the system. If populations are subjected to similar selective forces it is likely that they could find similar solutions, thus converging to similar phenotypes. Rocky headlands and sand dunes, usually occurring next to each other, are characterized by contrasting conditions at different localities. The headland cliffs are characterized by rocky soils rich in nutrients, are exposed to strong winds, and are constantly sprayed with salty water. Sand dunes are characterized by granular soils and loose matrix, they are poor in nutrients, drain water easily, and are prone to over heating. Headland populations are short, highly branched and with small and succulent leaves. Dunes are tall, with few branches and larger leaves. Experiments under controlled conditions revealed that these differences in morphology are genetically based, and that growth habit has evolved independently multiple times (Roda *et al.*, 2013).

Overall, the coastal populations of *Senecio lautus* are an ideal system to study speciation and adaptation. Although the parallel evolution of phenotypes is not uncommon, only few studies provide evidence for it, while knowledge on ecological speciation in plants appear particularly limited Thus, the parallel evolution of phenotypes *S. lautus* provide a unique opportunity to study ecological speciation, and also to investigate the influence that other forces could have on the evolution of RI –for example, the effect of genetic drift and recombination over varying geographic distances. *S. lautus* also meets many characteristics that can facilitate the joint study of ecology, genetics and evolution in the study of speciation. *S. lautus* are relatively short life cycle (annuals and biennials)(Ali, 1968), where a single flower can produce more than 50 seeds –allowing to create large experimental populations. This species displays strong self-incompatibility (Ornduff, 1964), favouring genetic studies by performing controlled crosses. Furthermore, in our lab genetic extensive genetic resources are developed and a linkage map has been recently constructed. In the following chapters, I provide evidence for the thesis that *S. lautus* evolved by natural selection and it is as a solid model for the study of the origin of new species.

References

Agrawal AF, Feder JL, Nosil P. 2011. Ecological divergence and the origins of intrinsic postmating isolation with gene flow. *International Journal of Ecology* **2011**.

Ali S. 1966. *Senecio lautus* complex in Australia. III. The genetic system. *Australian Journal of Botany* **14**: 317-327.

Ali S. 1968. *Senecio lautus* complex in Australia. IV. The biology of the complex. *Phyton (Horn, Austria)* **13**: 53-62.

Ali SI. 1964. *Senecio lautus* complex in Australia. I. Taxonomic considerations and discussion of some of the related taxa from New Zealand. *Australian Journal of Botany* **12**: 282-291.

Ali SI. 1969. *Senecio lautus* complex in Australia. V. Taxonomic interpretations. *Australian Journal of Botany* **17**: 161-176.

Berner D, Grandchamp AC, Hendry AP. 2009. Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. *Evolution* **63**: 1740-1753.

Bolnick DI, Fitzpatrick BM. 2007. Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics* **38**: 459-487.

Butlin RK. 2005. Recombination and speciation. *Molecular Ecology* **14**: 2621-2635.

Butlin RK, Galindo J, Grahame JW. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 2997-3007.

Clausen J. 1951. Stages in the evolution of plant species. *Stages in the evolution of plant species.*(6d).

Coyne. 1992. Genetics and speciation. *Nature* **355**: 511-515.

Coyne JA, Orr HA. 1997. " Patterns of speciation in *Drosophila*" revisited. *Evolution* **51**: 295-303.

Coyne JA, Orr HA. 2004. *Speciation*: Sinauer Associates Sunderland, MA.

Dobzhansky T. 1940. Speciation as a stage in evolutionary divergence. *American Naturalist*: 312-321.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**: 342-350.

Feder JL, Flaxman SM, Egan S, Comeault AA, Nosil P. 2013. Geographic mode of speciation and Genomic Divergence. *Annual Review of Ecology, Evolution, and Systematics* **44**: null.

Felsenstein J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**: 124-138.

Foster SA, McKinnon GE, Steane DA, Potts BM, Vaillancourt RE. 2007. Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. *New Phytologist* **175**(2): 370-380.

Fuller R, McGhee K, Schrader M. 2007. Speciation in killifish and the role of salt tolerance. *Journal of Evolutionary Biology* **20**: 1962-1975.

Gavrilets S. 2003. Perspective: models of speciation: what have we learned in 40 years? *Evolution* **57**: 2197-2215.

Gavrilets S. 2004. Fitness landscapes and the origin of species. *Austral Ecology* **30**: 610-611.

Hatfield T, Schluter D. 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* **53**: 866-873.

Hendry AP. 2009. Ecological speciation! Or the lack thereof? This Perspective is based on the author's JC Stevenson Memorial Lecture delivered at the Canadian Conference for Fisheries Research in Halifax, Nova Scotia, January 2008. *Canadian Journal of Fisheries and Aquatic Sciences* **66**: 1383-1398.

Hey J. 2006. Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development* **16**: 592-596.

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**: 747-760.

Jiggins CD, Naisbit RE, Coe RL, Mallet J. 2001. Reproductive isolation caused by colour pattern mimicry. *Nature* **411**: 302-305.

Lachaise D, David JR, Lemeunier F, Tsacas L, Ashburner M. 1986. The reproductive relationships of Drosophila sechellia with *D. mauritiana, D. simulans,* and *D. melanogaster* from the Afrotropical region. *Evolution*: 262-271.

Leinonen T, Herczeg G, Cano JM, Merilä J. 2011. Predation-imposed selection of threespine stickleback (*Gasterosteus aculeatus*) morphology: a test of the refuge use hypothesis. *Evolution*.

Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH. 2008a. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 3009-3021.

Lowry DB, Rockwood RC, Willis JH. 2008b. Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution* **62**: 2196-2214.

Machado CA, Hey J. 2003. The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**: 1193-1202.

Mallet J, Meyer A, Nosil P, Feder JL. 2009. Space, sympatry and speciation. *Journal of Evolutionary Biology* **22**: 2332-2341.

Mayr E. 1963. Animal species and evolution. *Animal species and their evolution.*

Muir G, Schloetterer C. 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus spp*.). *Molecular Ecology* **14**: 549-561.

Niemiller ML, Fitzpatrick BM, Miller BT. 2008. Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: Gyrinophilus) inferred from gene genealogies. *Molecular Ecology* **17**: 2258-2275.

Nosil P. 2004. Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**: 1521-1528.

Nosil P. 2008. Speciation with gene flow could be common. *Molecular Ecology* **17**: 2103-2106.

Nosil P. 2012. *Ecological speciation*: Oxford University Press.

Nosil P, Egan S, Funk D. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes:" Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**: 316-336.

Nosil P, Funk D, Ortiz-Barrientos D. 2009a. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**: 375-402.

Nosil P, Harmon LJ, Seehausen O. 2009b. Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* **24**: 145-156.

Nosil P, Vines T, Funk D. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**: 705-719.

Ornduff R. 1964. Evolutionary pathways of the *Senecio lautus* alliance in New Zealand and Australia. *Evolution*: 349-360.

Ostevik KL, Moyers BT, Owens GL, Rieseberg LH. 2012. Parallel ecological speciation in plants? *International Journal of Ecology* **2012**.

Presgraves DC. 2002. Patterns of postzygotic isolation in Lepidoptera. *Evolution* **56**: 1168-1183.

Radford I, Cousens R, Michael P. 2004. Morphological and genetic variation in the *Senecio pinnatifolius* complex: are variants worthy of taxonomic recognition? *Australian Systematic Botany* **17**: 29-48.

Ramsey J, Bradshaw Jr H, Schemske D. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**: 1520-1534.

Rieseberg LH, Desrochers AM, Youn SJ. 1995. Interspecific pollen competition as a reproductive barrier between sympatric species of Helianthus (Asteraceae). *American Journal of Botany*: 515-519.

Roda F, Ambrose L, Walter GM, Liu HL, Schaul A, Lowe A, Pelser PB, Prentis P, Rieseberg LH, Ortiz-Barrientos D. 2013. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Molecular Ecology* **22**: 2941-2952.

Rundle HD. 2002. A test of ecologically dependent postmating isolation between sympatric sticklebacks. *Evolution* **56**: 322-329.

Rundle HD, Nosil P. 2005. Ecological speciation. *Ecology Letters* **8**(3): 336-352.

Rundle HD, Whitlock MC. 2001. A genetic interpretation of ecologically dependent isolation. *Evolution* **55**: 198-201.

Schemske DW. 2010. Adaptation and the origin of species. *American Naturalist* **176**: S4-S25.

Schluter D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* **16**: 372-380.

Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* **323**: 737.

Sobel J, Chen G, Watt L, Schemske D. 2010. The biology of speciation. *Evolution* **64**: 295-315.

Sturtevant A. 1920. Genetic studies on *Drosophila simulans.* I. Introduction. Hybrids with *Drosophila melanogaster*. *Genetics* **5**: 488.

Thibert-Plante X, Hendry A. 2009. Five questions on ecological speciation addressed with individual-based simulations. *Journal of Evolutionary Biology* **22**: 109-123.

Thompson I. 2005. Taxonomic studies of Australian *Senecio* (Asteraceae): 5. The *S. pinnatifolius/S. lautus* complex. *Muelleria* **21**: 23-76.

Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology* **3**: 1572.

Via S, Bouck A, Skillman S. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* **54**: 1626-1637.

Figures

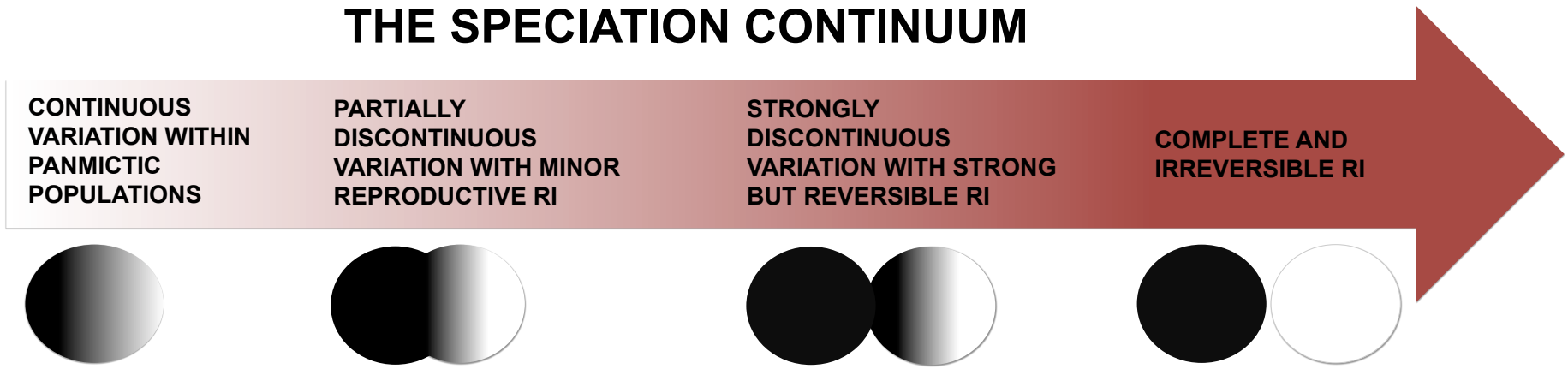

Fig. 1.1. The speciation continuum (modified from Hendry *et al.* 2009 and Lowry, 2012).

CHAPTER II

STRONG EXTRINSIC REPRODUCTIVE ISOLATION BETWEEN PARAPATRIC POPULATIONS OF AN AUSTRALIAN GROUNDSEL

ABSTRACT

Speciation with gene flow, or the evolution of reproductive isolation between interbreeding populations remains a controversial problem in evolution. This is because gene flow erodes the adaptive differences that selection creates between populations. Here, we use a combination of common garden experiments in the field and in the glasshouse to investigate what ecological and genetic mechanisms prevent gene flow and maintain morphological and genetic differentiation between coastal parapatric populations of the Australian groundsel *Senecio lautus*. We discovered that in each habitat extrinsic reproductive barriers prevented gene flow, whereas intrinsic barriers in F1 hybrids were weak. In the field, herbivores played a major role in preventing gene flow, but glasshouse experiments demonstrated that soil type also created variable selective pressures both locally and on a greater geographic scale. Our experimental results demonstrate that interfertile plant populations adapting to contrasting environments may diverge as a consequence of concurrent natural selection acting against migrants and hybrids through multiple mechanisms. These results provide novel insights into the consequences of local adaptation in the origin of strong barriers to gene flow in plants, and suggest that herbivory may play an important role in the early stages of plant speciation.

Key words: Parapatric, ecological divergence, speciation, extrinsic reproductive isolation, intrinsic reproductive isolation, prezygotic isolation, postzygotic isolation, *Senecio lautus,* predation.

INTRODUCTION

Local adaptation is thought to be a major contributor to the evolution of reproductive isolation (RI) between parapatric populations (Rundle & Nosil, 2005; Schemske, 2010). Although gene flow usually turns the odds against the formation of new species (Felsenstein, 1981), ecological contributions to RI in parapatry are not uncommon, having a long history in natural and experimental settings. For instance, adjacent populations of grasses on soils containing high or low levels of toxic heavy metals have evolved into morphologically and physiologically differentiated populations that persist despite gene flow (McNeilly & Antonovics, 1968; Antonovics & Bradshaw, 1970; Antonovics, 2006). Similarly, *Anthoxanthum odoratum* populations exposed to different environmental conditions have evolved both morphological differences and RI since the inception of the Park Grass Experiment in 1856 (Davies & Snaydon, 1976; Silvertown *et* al., 2005). These empirical results echo those from theory, where the predicted conditions for parapatric speciation seem to be common in nature (e.g., isolation by distance between populations, and patchy and linear habitats along rivers and coasts; Gavrilets, 2000). However, studies of RI often fail to identify the agents of divergent natural selection or quantify the relative contributions of multiple reproductive barriers to gene flow during the lifetime of organisms (c.f. studies reviewed in Lowry *et al.*, 2008a). As a consequence, our knowledge of which barriers trigger the speciation process, and the relative importance of extrinsic versus intrinsic barriers to gene flow remains limited in most studies of ecotype and species formation in plants.

Local adaptation creates barriers to gene flow in parapatry through various mechanisms (Schluter, 2001; Rundle & Nosil, 2005; Hendry *et al.*, 2007). When locally adapted populations exchange migrants, theory predicts that they will fare poorly in the environment of the sister population. This creates greater opportunities for interbreeding within rather than between populations, in turn limiting gene flow (Nagy & Rice, 1997; Hendry, 2004; Nosil, 2004; Thibert- Plante & Hendry, 2009). Local adaptation can also cause ecologically dependent reductions in F1 hybrid fitness, a phenomenon known as extrinsic postzygotic RI (Schluter, 2000; Rundle & Whitlock, 2001; Rundle & Nosil, 2005). Generally, ecologically dependent reductions in hybrid fitness occur because hybrids express intermediate parental phenotypic values for locally adapted traits (Barton & Hewitt, 1985; Schluter, 2000; Rundle & Nosil, 2005), thus rendering them unfit in parental habitats. The extent to which hybrids are ecologically disadvantaged depends on the form of inheritance for the traits under divergent natural selection (e.g., dominance versus additivity; Arnold, 1997; Barton, 2001; Berner *et al.*, 2011). However, reductions in F1 hybrid fitness could also result from their failure to cope with stressful conditions, including those experienced in the field (Coyne & Orr, 2004). This form of hybrid failure is also extrinsic but not specific to the environment of the parents

that produced the hybrid offspring. These mechanisms of postzygotic RI only manifest under field or stressful conditions but dissipate under controlled or benign conditions, such as those found in glasshouses (Hoffmann & Merilä, 1999; Bordenstein & Drapeau, 2001).

The mechanisms creating divergent natural selection are often difficult to identify and quantify. However, some habitat differences are known to contribute to local adaptation in plants (Kruckeberg, 1986; O'Dell & Rajakaruna, 2011). Edaphic and climatic differences usually cause strong divergent selection between populations, possibly because of drought (Stebbins, 1952; Bray, 2002), toxicity (Brady *et al.*, 2005), temperature (Keller & Seehausen, 2012) or a combination of them. These effects are common across many plant taxa (Kruckeberg, 1951; Wu *et al.*, 1975; Emms & Arnold, 1997; Nagy & Rice, 1997; Vekemans & Lefèbvre, 1997; Berglund *et al.*, 2004; Kay, 2006; Martin *et al.*, 2006; Sambatti & Rice, 2007; Lowry *et al.*, 2008a) and suggest that environmental stress could create strong extrinsic prezygotic and varying degrees of postzygotic extrinsic RI, thus playing an important role during the early stages of ecotype and species formation (Lowry, 2012).

Plant and animal interactions can also contribute to the evolution of RI between plant populations. The most famous examples involve systems where pollinators discriminate floral differences between ecotypes or species (Emms & Arnold, 1997; Bradshaw & Schemske, 2003; Kephart & Theiss, 2004; DellíOlivo *et al.*, 2011). A less studied biotic cause of RI is the contributions of herbivores and parasites to reductions in gene flow between populations (Sork *et al.*, 1993; Combes, 1996; Fritz *et al.*, 1999; Elias *et al.*, 2012). For instance, hybrids between populations of willows show differential responses to aphids and mites (Fritz *et al.*, 1994; Czesak *et al.*, 2004), similar to the adult hybrids of *Oenanthe conioides* and *O. aquatica* that are preferentially grazed by waterfowls and snails in the environment of *O. conioides* (Westberg *et al.*, 2010). Whether this kind of interaction between plants and invertebrates is important for the progress towards speciation and the origins of RI in plants requires further exploration.

Populations of the groundsel *Senecio lautus* that inhabit the sand dunes (Dune populations) and rocky headlands (Headland populations) along the Australian coast are an excellent system to study the origin and maintenance of ecotypes and the early stages of speciation. Often found adjacent to each other along the coast, Dune and Headland populations show marked morphological differentiation (Thompson, 2005; Fig. S2.1 for typical ecotype morphologies), which has evolved repeatedly and independently multiple times, and in the face of gene flow (Roda *et al.*, al. 2013a). Dune and Headland populations retain their morphologies in glasshouse conditions (see Abbott, 1976 for a similar case in European *Senecio*), and reportedly exhibit weak intrinsic reproductive

barriers (Ali, 1964; Ali, 1968). Previous transplant experiments in *S. lautus* suggest that some populations are adapted to their local environment (Radford *et al.*, 2004). However, little is known about what causes differentiation in the system and whether local adaptation is driving the evolution of RI in parapatric populations. We chose one of these parapatric pairs to investigate the evolution of reproductive isolating barriers in response to adaptation to contrasting coastal environments. Through field observations and common garden experiments in the field and the glasshouse we estimated various components of RI between these two parapatric populations, and identified possible ecological mechanisms causing divergent natural selection. We discuss how these results inform us about the relative contributions of extrinsic and intrinsic reproductive barriers during the early stages of speciation with gene flow.

MATERIALS AND METHODS

The system

Senecio lautus is a diverse complex of groundsels inhabiting a wide variety of environments in the South Pacific. Populations in Australia occupy diverse habitats including alpine and tablelands, woodlands, and coastal sand dunes and cliffs (Ali, 1964; Radford *et al.*, 2004; Thompson, 2005). Most populations exhibit strong self-incompatibility (Ornduff, 1964; and personal observations), and display variable life history traits ranging from annual and biennials to short- and long-lived perennials (Ali, 1968). Traits such as leaf morphology and plant architecture show strong associations with the environment in which populations are found, suggesting that the system consists of multiple ecotypes distributed across geography (Radford *et al.*, 2004; Roda *et al.*, 2013a). Populations inhabiting sand dunes (Dune ecotype) and rocky headlands (Headland ecotype) are related by geography and not habitat, and each pair is genetically differentiated from such other pairs (Roda *et al.*, 2013a). The pair at Cabarita Beach displays strong heterogeneous divergence across its genome and moderate average Fst (0.04) compared to comparisons between allopatric populations (Fst 0.12-0.2; Roda *et al.*, 2013a). We studied two coastal forms of *Senecio lautus* that grow on the sand dune (S28° 19' 54.66" E153° 34' 17.04") and rocky headland (S28° 21' 45.07" E153° 34' 46.82") environments at Cabarita Beach, in northern New South Wales, Australia (Fig. S2.1). Although the beach and headland are abutting environments, a small town separates the Dune and Headland populations by ~800m. The beach and parts of the headland remain connected via rocky outcrops and small cliffs. Of note, unpublished data suggest that these populations and similar pairs, are at least 30,000 years old, whereas towns in Australia are less than 200 years old, suggesting that effects from urbanization on population dynamics are recent compared to those that

drove past evolutionary divergence. Usually only a few meters separate other population pairs along the coast where towns do not interrupt the shore (e.g., the population pair at Coffs Harbour, NSW are separated by 3m). The Headland habitat is characterized by rocky mineral-rich soil, exposed to constant salt spray and strong winds. In this habitat individuals exhibit a compact architecture, being short and matt forming (prostrate), heavily branched, with small succulent leaves. In the Dune habitat, the soil is sandy, poor in nutrients, and susceptible to heating during sunny days. Individuals from the sand dunes are tall, have few branches with large and thin leaves (Radford *et al.*, 2004; Thompson, 2005; Fig. S2.1).

Crosses

Seeds were collected from 30 individuals from each coastal population of Cabarita Beach in 2009. Laboratory seed stocks for parental and F1 hybrids full-sib families –derived from randomized crosses with equal contribution between parental types– were created in glasshouses at the University of Queensland, St Lucia, QLD, Australia. Scarified seeds (1mm trimmed at the micropyle side) from each family were germinated on moist filter paper in petri dishes. Seeds were kept in dark and controlled conditions for three days to induce root elongation, and subsequently placed under the light for 7 days to induce vegetative growth. One week old seedlings were transplanted into 2.5L pots filled with standard potting mix and transferred to a glasshouse with constant temperature (25°C) and 12h:12h light:dark cycle. Flowering time was recorded, and plant morphology measurements were made on adult individuals with flowers (~2 months after germination). Intra and interpopulation crosses were performed twice a day by gently rubbing flower heads (capitula): each flower head was crossed at least three times to maximize the number of florets producing seeds. We kept track of unpollinated flowers to check for self-pollination, but did not find a single seed resulting from self-pollination.

Intrinsic RI

We calculated seed set by estimating the proportion of fertilized seeds in flower heads (Fig. S2.2). We divided the number of fertilized seeds in an interpopulation cross by the average number of seeds produced in parental intrapopulation crosses (Coyne & Orr, 2004). We calculated F1 seed set for two other parapatric population pairs (Hat head and Lennox Head) and two allopatric comparisons (Cabarita beach versus Byron Bay and Lamington National Park versus Port Macquarie) to provide context for our focal result (see Table S1 for the geographic location of all populations). The Tableland population from Lamington National Park, another member of the *S. lautus* species complex, is a perennial ecotype found far from the coast next to the edges of tropical

forests. It can reach several meters in height, and in contrast to coastal types, has long and ovate serrated leaves (Radford *et al.,* 2004; Roda *et al.,* 2013).

Flowering time differences between ecotypes

To investigate if the natural populations are phenologically separated, we counted the number of flowers per individual each month between February 2011 and January 2012. We report the total number of flowers per month in the population divided by the total number of individuals to control for changes in population size during the year. Because Cabarita Beach sand dunes are linear and narrow, we sampled flowers in individuals present in the same 80m x 4m transect. On the Headland population we sampled individuals in three different rocky grooves where the majority of individuals resided. We could not sample a few individuals growing on the cliffs due to high risk of falling.

Extrinsic RI

Soil experiments

We performed reciprocal germination and establishment experiments under controlled conditions in the glasshouse using soil collected in each locality. We filled two sets of eight plastic trays with fresh soil from either the sand dunes or the rocky headlands at Cabarita Beach. Seeds belonging to 20 families of each Dune and Headland ecotype were sown into 16 trays (each family was represented by one seed per tray) for a total of 640 seeds. Seeds were sown on top of the soil in a fully randomized design within each tray. Soil was sprayed with water three times a day to keep it moist, and tray position on the shelf was switched daily. Room temperature was at 25°C through a 12h:12h light:dark cycle. Seed germination and death occurred during the first 19 days of the experiment. A nominal logistic mixed model (GLM, Generalized Linear Model for binary data) was used to analyze the proportion of germination for each cross type (including parental and hybrid crosses) within soil, using the *lmer* package in R (R Core Team, 2012). The model included cross type as a fixed effect, and tray and family as random effects.

To further study the germination of Dune and Headland ecotypes from Cabarita Beach at other localities where Dune and Headland populations also grow, we conducted a second soil experiment. We used soil from three different sites (two in addition to Cabarita Beach) collected from the sand dunes and rocky headlands at Lennox Head and Stradbroke Island. We filled 2 trays per locality for a total of 6 trays. Using the same set up conditions as the experiment above, we sowed a total of 420 seeds (35 seeds for 5 families of each ecotype per tray). Germination and mortality were recorded daily during the first 33 days after which no more germination or death occurred.

Statistical analyses were performed using a GLM to test for the full interaction among soil, locality and cross type. We also performed an analysis within soil to test for germination differences between ecotypes. Cross type and soil were treated as fixed effects, and tray and family as random effects.

Reciprocal transplant experiment in the field

A total of 1760 seeds belonging to 70 families of Dune, 70 of Headland and 80 families of reciprocal F1s (F1-D and F1-H correspond to F1 individuals with either a Dune or Headland mother, respectively) were sown directly into the field in four plots in both the Dune and Headland environments (one seed per family per plot in a fully randomized design) in November 2010. To track sees in the field, they were individually pasted at their mid point onto toothpicks using Selleys Parfix superglue (ensuring neither extreme of the seeds was covered that would have obstructed germination). Control experiments in the laboratory and other reciprocal transplant experiments (not shown here) have shown that gluing seeds to toothpicks does not affect germination rates. Seeds were sown 5mm beneath the soil surface. Plots were covered with a 50% UV protection mesh (HDPE UV stabilized forest green exterior fabric) to prevent loss of seeds during the rainy season. Seeds were lightly sprayed with water once a day for the first two weeks of the experiment to keep the surface of the soil moist. Experiments were held during the wet season and an average of 313.69mm of rainfall fell during the first 45 days of the experiment.

Germination and mortality were recorded twice a day for the first month. Seedlings were considered killed by herbivores if only their stalks were found after two consecutive survivorship measurements. We termed predators invertebrates that were found on seedlings and ate both cotyledons leading to plant death. The few plants that disappeared between two consecutive measurements were also considered killed by predators. Analyses with and without these plants did not change the interpretation of results. Plants that wilted and died slowly were considered to have died by “other causes”, likely involving drought and predation. Because these individuals died slowly (i.e., progressive desiccation of the seedling), we were able to monitor death progression over several survivorship measurements. Survivorship data were taken over several hours during the morning and afternoon of each day of the first month, when most individuals died. For the following 4 months we visited the site weekly, and for the rest of the experiment every second week. A GLM was used to analyze the proportion of either germination or death due to predation for crosses within environment; the model included cross type as a fixed effect, and family and block as random effects. To further evaluate the effects of cross type on total survivorship, we performed a GLM using the Poisson distribution (to account for the large proportion of individuals

that died in the two environments) on the number of days alive. We censored the plants up to when the first flower was produced in each habitat (sand dunes day 275 and rocky headland day 149). Standard survivorship analyses (Fox, 2001) did not affect the conclusions derived from our results). Finally, we conducted a GLM to test the effect of parental and hybrid genotypes on the average number of flowers in each environment. We fitted general linear models using restricted maximum likelihood in the *lmer* package in R.

Strength of RI

We calculated the strength of several intrinsic (*I*) and extrinsic (*E*) RI barriers between the two coastal ecotypes, following the approach by Lowry *et al.*, (2008a). We calculated the following reproductive barriers: (1) *Flowering asynchrony in the field*, as RI_{phen1}=1-(observed/expected interpopulation matings)/(observed/expected intrapopulation matings). (2) *Immigrant inviability* (*E*), or whether migrant seeds had difficulties establishing in the alternative ecotype environment, as $RI_{imm}=1-(w_i/w_n)$, where w_i is the mean number of surviving migrant individuals, and w_n the surviving local type. (3) *Hybrid viability* (*I*, *E*), or whether hybrid seedlings germinated and survived equally well as their parents in field or controlled conditions, as $H_{hf}=1-(v_{mean}F1/v_{local})$, where v_{mean} is the average survivorship for F1 hybrids, and v_{local} is the average survivorship of the local ecotype. In the field, we also estimated hybrid viability by only taking into account mortality of individuals that germinated, thus disentangling the effects of lack of hybrid seed germination from hybrid mortality. (4) *Hybrid seed set* (under controlled conditions) (*I*) or whether the proportion of fertilized seeds in a flower head from an inter-ecotype cross differed from an intra-ecotype cross $RI_{seed\ set}=1-(Pf_{inter}/Pf_{intra})$ were Pf stands for proportion fertilized. We estimated the total cumulative RI for each ecotype taking into account the absolute contribution of each reproductive barrier in the study according to the methodology in Lowry *et al.*, (2008a). Finally, we calculated the magnitude of local adaptation (local adaptation index) for each of the crosses as described in Hereford, (2009). Three estimates of local adaptation were obtained depending on the fitness measure taken into account: average number of days in which each cross type was alive until the end of the experiment, average number of flowers per individual, and the product of these two variables.

RESULTS

Ecotypic differences

In Cabarita Beach, Dune individuals (D) were always erect or decumbent, whereas all Headland individuals (H) were short and prostrate. In the glasshouse, Headland individuals –with one exception– were prostrate, and all Dune individuals were erect or decumbent ($F_{1,37}$=33.5013, p<0.0001). The Headland population also had individuals with more branches than the Dune population ($F_{1,37}$=28.6556, p<0.0001). In the field, the Dune population flowered little or did not flower from November to March, while the Headland population flowered throughout the entire year (Fig. 2.1). In the glasshouse, Headland plants flowered after 8 weeks, while Dune individuals flowered after 10 weeks. In the glasshouse, both populations flowered for~4 months, after which plants stopped producing new leaves or flowers. Consistent with previous reports for the system (Ali, 1968; Radford *et al.*, 2004), growth habit differences between Cabarita Beach populations were retained in the glasshouse, however flowering time was affected by the environment in which they grew.

Intrinsic RI

Seeds produced in the glasshouse from crosses within or between populations were highly viable (germination success in moist filter paper > 98% for both parents and hybrids). Hybrid and parental seed set was similar (mean seed set: H=0.45 ±0.08; D=0.48±0.08, F1=0.49±0.07, $F_{2,62}$=0.0825, p=0.9358) indicating that intrinsic RI is weak between the Dune and Headland population at Cabarita Beach. Intrinsic RI in other population pairs was generally weak (Table 1). The population pair at Hat Head showed the greatest level of RI, but only in one direction of the cross. We found negative values of RI in some crosses, particularly in the cross between a Tableland and a Headland population. Overall, F1 hybrid fitness is similar to parental fitness in the *S. lautus* populations studied here.

Extrinsic RI

Soil experiments in the glasshouse

Although overall germination success in the glasshouse was low, parental seeds germinated equally well in both sandy and rocky soils from Cabarita Beach (D soil, z=-1.732, p=0.0834; H soil, z=-1.059, p=0.290; Fig. S2.3, Table S2.2). Headland and Dune seeds sown in soil collected from other dune and headland localities showed variable patterns of germination success (interaction model, X^2=36.807, p<0.0001; Fig. 2, Table S2.2). Germination differences were most pronounced in Lennox Head soil (D soil, z=-2.794 p=0.005; H soil, z=2.942, p=0.003) but less apparent and

asymmetric in Stradbroke Island soil (D soil, z=-1.938, p=0.053; H soil, z=-0.549 p=0.583). Patterns of germination success did not change in Cabarita Beach soil (D soil, z=-0.804, p=0.421; H soil, z=-0.305, p=0.760; note that this is a replicate of the experiment above).

Reciprocal transplant experiments in the field

Germination and mortality: Overall germination success in the field was low (Fig. 3a, Table S2.2), particularly in the Headland environment (z=-9.279, p<0.001), but did not differ between parental crosses in either of the two habitats (sand dunes, z=-0.913, p=0.361, rocky headland, z=-1.673, p=0.0943). In contrast, F1 hybrids germinated with significantly greater success than either parent in the field, particularly in the sand dunes (z=3.137, p=0.0017), and when they carried a Dune cytoplasm (z=3.973, p<0.001). Germination occurred in two independent bouts, but this did not affect the overall pattern of germination success between crosses and in either environment (Table 2.2). Censored adult survivorship differed between crosses in both environments (Fig. 4 and Table S2.3). In the sand dunes, D families survived better than other crosses (H, z=-5.847, p<0.001; F1-H, z=-3.158, p=0.0015 and F1-D, z=-3.035, p=0.0024). In the rocky headland, where selection was strongest, H families showed the highest survivorship, although differences were not significant amongst cross types (D, z=-1.396, p=0.1630; F1-D, z=0.459, p=0.6460 and F1-H, p=0.7820). Overall, the combined effects of germination and mortality indicate that the demography of transplanted populations is qualitatively different between environments (Fig. 2.3b). Thus, when the proportion of individuals that had germinated and survived was plotted throughout the course of the experiment (i.e., until most of the population died; Fig. 2.3b) it was evident that (i) germinated seed of all types was subject to mortality, although absolute mortality was greater for F1s, but intermediate between the two parental types, (ii) the local type (D or H in its local environment) performed better across the duration of the experiment (see fecundity results below), (iii) by the end of the experiment (~500 days) a proportion of each type of seed had germinated and produced individuals that had completed their life cycle (Fig. 2.3b), and (iv) the total number of individuals alive at the first day of flowering (see grey vertical line in each panel Fig 2.3b) in each environment differed amongst cross type and in the direction of local adaptation (X^2=16.3, df=3, p=0.001; Table 2.3).

Mortality due to herbivory: We found herbivores on the seedlings planted at Cabarita Beach (e.g., Fig. S2.4). Because herbivores killed plants, we refer to this process as predation. Death due to predation occurred only when seedlings had green cotyledons (once seedlings produced new leaves, we detected mortality events due to other causes, possibly drought). In both habitats, herbivores killed immigrants more often than local parents (sand dunes, z=3.753, p=0.0001; rocky headland,

z=2.159, p=0.0309; Fig. 5, Table S2.3). In both environments, hybrids were attacked significantly more than the local type, with the Headland cytoplasm conferring a slight advantage in the rocky headland (sand dunes, F1-D, z=3.763, p=0.0002; F1-H, z=2.904, p=0.0037; rocky headland, F1-D, z=2.124, p=0.0337, F1-H, z=0.766, p=0.4434; Fig. 2.5, Table S2.3 for proportions of predated individuals). Crosses did not show differences in mortality due to "other possible causes" of death in any of the two environments (sand dunes, H, z=0.235, p=0.8139; F1-D, z=-0.404, p=0.6863; F1-H, z=-0.017, p=0.9863; rocky headland, D, z=0.103, p=0.9181; F1-D, z=-0.038 p=0.9699, F1-H, z=-0.186, p=0.8526).

Number of flowers: The first flower head buds on Dune plants appeared after 275 days in the sand dunes and after 191 days in the Headland. In contrast, Headland individuals never flowered in the sand dunes, but flowered after 191 days in the rocky headland. The average number of flowers heads per individual (including those that did not flower) was randomly distributed with respect to cross type (sand dunes, $F_{3,79}$=0.779, p=0.5092, rocky headland, $F_{3,42}$=0.347, p=0.7912; Table 2.3).

Please refer to Table S2.2 and S2.3 for means and S.E. for all fitness components measured in each experiment, and see Table S2.4 for a combined summary of all statistical tests performed.

Strength of RI

For the barriers we measured, cumulative RI for Cabarita Beach populations was 0.88 in the sand dunes, with average prezygotic and postzygotic strength of 0.59 and 0.11, respectively. In the headland, cumulative RI was 0.76, with average prezygotic and postzygotic strength of 0.55 and 0.04, respectively. Overall, intrinsic reproductive barriers were absent (Table 2.1), whereas extrinsic reproductive barriers were strong (Table 2.4). In particular immigrant inviability and extrinsic postzygotic isolation seemed to play a major role in preventing gene flow between the two populations, although asymmetrically. Positive indexes of local adaptation, as measured by Hereford, (2009) showed that –when viability was the measure of fitness– each ecotype performed best in its own environment and F1 performance was intermediate but slightly better depending on whether they carried the local cytoplasm (Table 2.5, and S2.4 for further details). This was not the case when fitness measurements included number of flowers, although these measurements relied on small population sizes during the flowering season.

DISCUSSION

We have shown that extrinsic reproductive isolation (RI) creates strong barriers to gene flow between two neighboring coastal populations of the Australian groundsel *Senecio lautus*. These populations show morphological differences with a strong genetic basis that persist despite constraints of gene flow on divergence (Roda *et al.*, 2013a). We discovered that predators were the main cause of divergent natural selection and they contributed to selection against both migrants and hybrids between parental populations. Finally, and different to several previous studies connecting local adaptation and the origins of RI in plants (reviewed in Lowry, 2008a), we found that although F1 hybrids germinated and established initially much more than parents, they eventually suffered more individual losses to predators and other causes (possibly drought) than their parents. The fact that F1 hybrids showed hybrid vigor during the initial stages of development, suggests that heterosis and development may interact during the evolution of extrinsic RI. Below we discuss these main findings and their implications for understanding the progress towards speciation and the origins of RI in plants.

Lack of intrinsic RI in the F1 generation

We did not detect intrinsic RI between the two populations at Cabarita Beach, consistent with our measures of F1 seed set between allopatric pairs (Table 2.1) and with previous reports in the system where most populations were easily crossed and pollen fertility was generally high (Ornduff, 1964). Nevertheless, it is possible that hybrid fitness is reduced in later generations (i.e., hybrid breakdown), a common phenomenon detected in many plants (e.g., Fishman & Willis, 2001; Moyle & Graham, 2005). Although we did not measure F2 hybrid fitness between Cabarita Beach populations, we have unpublished data from other *S. lautus* populations where we have found hybrid breakdown, possibly suggesting that Dobzhansky-Muller incompatibilities (e.g., dominant x recessive, and recessive x recessive) have accumulated in the system. It is also possible that other intrinsic barriers such as conspecific pollen precedence might play an important role in this system, particularly when insect communities are shared across coastal habitats (White, 2008; Fig. S2.5).

Absence of some forms of intrinsic RI between populations adapting to contrasting environments is not uncommon, and may be normal during the early stages of speciation (Schluter, 1998; Hendry *et al.*, 2007), although there are clear examples where the two evolve together (Macnair & Christie, 1983). For instance, species adapted to serpentine soil lack intrinsic RI, yet they persist in parapatry, and sometimes completely cease exchanging genes (Brady *et al.*, 2005; Harrison & Rajakaruna, 2011). Similarly, inland and coastal populations of *Mimulus guttatus* show weak intrinsic RI, yet

they are largely reproductively isolated from one another (Lowry *et al*., 2008b). Finally, invertebrate (Via *et al.*, 2000; Nosil, 2007; McBride & Singer, 2010) and vertebrate (Hatfield & Schluter, 1999; Fuller *et al.*, 2007) species have populations that display strong morphological differentiation but weak intrinsic RI. Whether extrinsic barriers to gene flow can be considered triggers of speciation requires further work (Nosil *et al*., 2009) but simulations have shown that strong selection against migrants is an effective and rapid way to reduce gene flow between populations (Hendry, 2004; Thibert-Plante & Hendry, 2009).

Heterotic and cytoplasmic effects on extrinsic RI

In our field experiments we found strong differences between parental and hybrid types in germination and survival, contrary to glasshouse results (Fig. 2.3). Germination success was significantly higher in hybrids compared to parental types, consistent with other studies where hybrids show heterotic effects (reviewed in Lowry *et al.*, 2008a; Sambatti *et al.*, 2012). These heterotic effects could result from 1) outcrossing breaking down homozygocity of detrimental recessive alleles within populations (inbreeding depression), or 2) hybridization creating loci of heterozygous advantage (overdominace, Charlesworth & Willis, 2009). Although we were unable to distinguish between these two possibilities, hybrids superiority was only found under field conditions and not in the glass house, where both parentals and hybrids performed similarly. In addition, this heterotic effect only during the early life cycle stages, suggesting an influence of parental cytoplasmic effects (i.e., the mother of the F1 hybrid had effects on its germination success in the field, Fig. 2.3 and 2.5, and see discussion below). Despite reasons for the initial extrinsic heterotic effect remain unclear, a release of antagonistic effects in hybrids (Burke & Arnold, 2001) might partially explain our observations: where genes controlling growth and reproduction no longer function together at later stages of development.

Our field experiments suggest that there could be extrinsic cytoplasmic effects on hybrid fitness. This result echoes those found on *Ipomopsis aggregata* and *I. tenuituba* (Campbell & Waser, 2001) and in *Chamaecrista fasciculata* (Galloway & Fenster, 1999) where a similar distinction in hybrid fitness between benign and field conditions were found. The reason for these effects is unknown, but it could be the result of stress dependent cyto-nuclear incompatibilities (Coyne & Orr, 2004). In our field experiment, F1 hybrids showed mortality patterns consistent with accumulation of stress through development: F1 individuals displayed hybrid vigor during early development but failed to survive to late developmental stages. However, most deaths were due to predation, so for stress to be a viable hypothesis, we must predict that predators preferred stressed to non-stressed hybrids. Although this possibility deserves further experimentation, two observations suggest that stress

alone is not the main cause: first, predators ate healthy plants, and not wilted plants. And second, previous studies have found that some phytophagous insects frequently prefer healthy plants over water stressed ones, as water stress can interfere with their ability to avail nitrogen (Huberty & Denno, 2004). Field experiments with cross types containing an increasing proportion of a local genome (e.g., reciprocal backcrosses, F1, and F2 hybrids) could help disentangle environmentally-dependent and intrinsic causes of mortality in this system (Rundle & Whitlock, 2001). Maternal provisioning of nutrients to seeds could also affect hybrid performance, but adding seed mass as a covariate in our analyses did not affect germination or survivorship (data not shown).

Predation creates extrinsic RI

Systems where immigrant inviability is strong are suggestive of strong extrinsic postzygotic RI (Lowry, *et al.*, 2008a). Although this has been demonstrated in a few cases in animals (Via *et al.*, 2000), such a link is less clear in plant systems (Lowry *et al.*, 2008a). For instance, in studies of the sister species *Ipomopsis agregata* and *I. tenuitubai*, selection against migrants was noticeable, but F1s performed on average similarly to parental types, although there were reductions in survival depending on the direction of the cross (Campbell & Waser, 2001). Similarly, in *Artemisia tridentata* subspecies, hybrids displayed a fitness advantage in most habitats in which they grew (Miglia *et al.*, 2005), thus possibly facilitating the opportunity of gene flow between subspecies. However, manipulative experiments in some plants have demonstrated that selection against migrants and hybrids could evolve quickly (Jain & Bradshaw, 1966; Davies & Snaydon, 1976; Hendry *et al.*, 2007).

In our experiments, herbivory created both extrinsic prezygotic and postzygotic RI barriers between parapatric populations of *S. lautus* (Table 2.4). Previous studies report that herbivores partially consumed fractions of leaves or flowers of adult individuals (e.g., Combes, 1996; Fritz *et al.*, 1999); however, our results are more related to animal examples where individuals were killed by attacks from other organisms (Langerhans *et al.*, 2007; Nosil, 2004) due to predation of newly emerged seedlings resulting in deaths. Although we cannot currently explain the causative agents for differential predation on parental versus migrant and hybrid seedlings, studies on other species in the *Senecio* genus have revealed that toxic alkaloids may serve as plant defenses against insects, and that production of such alkaloids is largely dependent on the environment where the species occurs (Kirk *et al.*, 2010). It would not be surprising if divergence in the type and amount of alkaloids in the cotyledons of *S. lautus* seedlings were responsible for the evolution of extrinsic RI between ecotypes. However, further studies on the genetics of immigrant and extrinsic postzygotic RI are required to further understand how they can evolve concurrently. Overall, our results link

proximate causes (predation) with ultimate causes of divergence (natural selection; Laland *et al.*, 2011), and suggest that natural selection reduces the exchange of genes between the Dune and the Headland ecotypes through the formation of maladapted hybrids and selection against migrants.

Although predation is an important agent of selection, other environmental variables could contribute to divergence between Dune and Headland populations. For instance, a previous study in the *Senecio* system found that soil composition differed drastically between dune and Headland environments, and with large localized variation (Roda *et al*., 2013b). Furthermore, a large proportion of allelic variation in Dune and Headland parapatric pairs, including Cabarita Beach, correlated with variation in abiotic elements found in their soils, including salt, metal, and nutrient content (Roda *et al*., 2013b). These results suggested that soil content contributes to adaptation to sand dunes and rocky headlands and may partially explain why we saw variable germination rates between the Dune and Headland environments (Fig. 2.2).

Investigating predation as a source of divergent natural selection

We discovered that predators may constitute a strong selective agent in *Senecio,* but further studies should examine this mechanism in more detail. Experiments including enclosures would allow estimating the survival of seedlings from the two ecotypes and their hybrids in the two environments in absence of potential invertebrate predators. Consistent with our results, we expect that populations in both environments would show similar survival. This prediction is only accurate for the earliest days of the life cycle of seedlings (~ one month after germination); time after which we cannot discard other environmental factors may contribute to create RI. We also require experiments aiming to identify the full range of invertebrates predating on the seedlings in each of the habitats. In our previous observations we found caterpillars from the moth genus *Spilosoma,* eating the whole seedlings in the rocky headlands. However, other invertebrates like small snails and ants were abundant in both environments. Finally, invertebrate's preference for food source should be explored. If local invertebrate communities feed more on migrant and hybrids, alternative hypothesis like stress can be discarded. Finally, if this last is proven, experiments that identify physiological and genetic differences between ecotypes should be conducted.

Stages of speciation

Studies of speciation have considered useful to treat ecotypes as a stage of the process of species formation (Clausen *et al.*, 1947; Lowry 2012). This view can help us understand the role of ecology on speciation and the relative contributions of reproductive isolating barriers to each stage of the process. The mechanisms facilitating transitions between ecotype and species remain mysterious, although it is possible that once extrinsic barriers to gene flow establish, neutral differentiation

accumulates, thus leading to the evolution of genetic incompatibilities responsible for various forms of intrinsic RI in a system (Nosil *et al*., 2008). Additionally, general extrinsic barriers to gene flow may facilitate novel sweeps in each population if the loci under selection are linked to those under initial divergent selection (Hendry *et al.,* 2007). This form of selection may be powerful and fast in creating further barriers to gene flow including intrinsic ones. Likewise, environmental effects on the time of reproduction (e.g., flowering time differences were marked in the natural populations in the field but limited in the glasshouse) may reduce gene flow even in the absence of genetic divergence between populations and thus promote the subsequent evolution of genetically based reproductive barriers (e.g., Thibert-Plante & Hendry, 2011). Finally, gamete competition within populations could lead to conspecific pollen precedence between populations, perhaps rapidly driving the evolution of strong barriers to gene flow during the early stages of speciation (Howard, 1999).

Overall, Cabarita Beach populations have evolved extrinsic barriers to gene flow, and the strength of RI is relatively high (see results and Table 2.4). According to the classification of stages in the progress toward speciation in Hendry *et al.* (2009) –that ranges from totally panmictic populations to completely and irreversibly isolated species– Dune and Headland populations at Cabarita Beach are ecotypes that seem to be in an intermediate stage of divergence: "Strongly discontinuous variation between populations with strong but reversible RI" (Hendry *et al.,* 2009). Whether the Dune and Headland populations will become discrete species is not currently possible to know, but our recent multilocus estimates of divergence between multiple coastal population pairs suggest that divergent natural selection can take populations to varying degrees of divergence in some cases with no detectable levels of gene flow from molecular markers (Melo *et al.,* unpublished results). Multiple Dune and Headland parapatric pairs in *S. lautus* could help us to identify the factors affecting (favoring/constraining) the progress toward ecological speciation and thus inform us about the different points at which the distinct pairs could be in the ecological speciation continuum (Hendry *et al.*, 2009; Nosil *et al.*, 2009). We expect that studies on the genetic basis of adaptation in this system will shed light as to whether regions responsible for extrinsic RI could persist.

CONCLUSIONS

Field and glasshouse experiments described here suggest that both immigrant inviability and extrinsic postzygotic isolation create strong barriers to gene flow between Dune and Headland populations of *S. lautus*. In agreement with other studies of diverging populations adapted to contrasting habitats (Clausen *et al.,* 1947; McNeilly & Antonovics, 1968; reviewed in Lowry *et al.*,

2008b), immigrant inviability was stronger than hybrid inviability (Table 4; Ramsey *et al.,* 2003, Nosil *et al.,* 2005), intrinsic barriers were barely noticeable in F1 hybrids, and natural selection (local adaptation index) was most effective before the reproductive season, in particular, at the seedling stage. In general, our results resemble those found in animal systems such as *Timema* walking sticks (Nosil, 2004), sticklebacks (Hatfield & Schluter, 1999), and pea aphids (Via *et al.*, 2000), but contrast with results found in other plant systems such as *Mimulus,* where F1 extrinsic postzygotic reproductive isolation between ecotypes is rather weak (Lowry *et al.*, 2008b), but strong when locus specific effects are considered in later hybrid generations (Lowry & Willis, 2010). We suggest that ecological reproductive isolation between plant and animals may follow similar paths.

Acknowledgements

This research was funded by the Australian Research Council grants to D.O-B. Alexander Gofton provided assistance in the field. Nick Barton, Mohamed Noor, Patrik Nosil, Andrew Hendry, Camilo Salazar, Diana Bernal, Huanle Liu, Maddie James, and Thomas Richards provided very useful feedback on previous versions of this manuscript. We are grateful to Richard Abbott for his editorial guidance, and for the useful comments that we received from three anonymous reviewers. Simon Bloomberg provided statistical advice. Antonia Posada provided help throughout the various steps of the experiments. We thank the Tweed Head Shire Council for allowing us to carry out our research at Cabarita Beach

References

Abbott RJ. 1976. Variation within common groundsel, *Senecio vulgaris L*. I. Genetic response to spatial variations of the environment. *New Phytologist* **76**: 153-164.

Ali SI. 1964. *Senecio lautus* complex in Australia. I. Taxonomic considerations and discussion of some of the related taxa from New Zealand. *Australian Journal of Botany* **12**: 282-291.

Ali S. 1968. *Senecio lautus* complex in Australia. IV. The biology of the complex. *Phyton (Horn, Austria)* **13**: 53-62.

Antonovics J. 2006. Evolution in closely adjacent plant populations X: long-term persistence of preRI at a mine boundary. *Heredity* **97**: 33-37.

Antonovics J, Bradshaw AD. 1970. Evolution in closely adjacent plant populations. VIII. Clinal patterns at a mine boundary. *Heredity* **25**: 349-362.

Arnold ML. 1997. *Natural hybridization and evolution*: Oxford University Press, USA.

Barton N. 2001. The role of hybridization in evolution. *Molecular Ecology* **10**: 551-568.

Barton NH, Hewitt G. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**: 113-148.

Berglund ABN, Dahlgren S, Westerbergh A. 2004. Evidence for parallel evolution and site-specific selection of serpentine tolerance in *Cerastium alpinum* during the colonization of Scandinavia. *New Phytologist* **161**: 199-209.

Berner D, Kaeuffer R, Grandchamp AC, Raeymaekers J, Räsänen K, Hendry A. 2011. Quantitative genetic inheritance of morphological divergence in a lake–stream stickleback ecotype pair: implications for reproductive isolation. *Journal of Evolutionary Biology* **24**: 1975-1983.

Bordenstein S, Drapeau M. 2001. Genotype- by- environment interaction and the Dobzhansky–Muller model of postzygotic isolation. *Journal of Evolutionary Biology* **14**: 490-501.

Bradshaw H, Schemske DW. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* **426**: 176-178.

Brady KU, Kruckeberg AR, Bradshaw Jr H. 2005. Evolutionary ecology of plant adaptation to serpentine soils. *Annual Review of Ecology, Evolution, and Systematics* **36**: 243-266.

Bray EA. 2002. Classification of genes differentially expressed during water- deficit stress in Arabidopsis thaliana: An analysis using microarray and differential expression data. *Annals of Botany* **89**: 803-811.

Burke JM, Arnold ML. 2001. Genetics and the fitness of hybrids. *Annual Review of Genetics* **35**: 31-52.

Campbell DR, Waser NM. 2001. Genotype-by-environment interactions and the fitness of plant hybrids in the wild. *Evolution* **55**: 669-676.

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. *Nature Reviews Genetics* **10**: 783-796.

Clausen J, Keck DD, Hiesey WM. 1947. Heredity of geographically and ecologically isolated races. *The American Naturalist* **81**: 114-133.

Combes C. 1996. Parasites, biodiversity and ecosystem stability. *Biodiversity and Conservation* **5**: 953-962.

Coyne JA, Orr HA. 2004. *Speciation*: Sinauer Associates Sunderland, MA.

Czesak M, Knee M, Gale R, Bodach S, Fritz R. 2004. Genetic architecture of resistance to aphids and mites in a willow hybrid system. *Heredity* **93**: 619-626.

Davies M, Snaydon R. 1976. Rapid population differentiation in a mosaic environment. *Heredity* **36**: 59-66.

DellíOlivo A, Hoballah ME, G¸bitz T, Kuhlemeier C. 2011. Isolation barriers between *Petunia axillaris* and *Petunia integrifolia* (Solanaceae). *Evolution* **65**: 1979-1991

Elias M, Faria R, Gompert Z, Hendry AP. 2012. Factors influencing progress toward ecological speciation. *International Journal of Ecology*. 2012: Article ID 235010.

Emms S, Arnold M. 1997. The effect of habitat on parental and hybrid fitness: transplant experiments with Louisiana irises. *Evolution* **51**: 1112-1119.

Felsenstein J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**: 124-138.

Fishman L, Willis JH. 2001. Evidence for Dobzhansky-Muller incompatibilities contributing to the sterility of hybrids between *Mimulus Gutatus* and *M. Nasutus. . Evolution* **55**: 1932-1942.

Fritz R, Moulia C, Newcombe G. 1999. Resistance of hybrid plants and animals to herbivores, pathogens, and parasites. *Annual Review of Ecology and Systematics*: **30**: 565-591.

Fritz R, Nichols-Orians C, Brunsfeld S. 1994. Interspecific hybridization of plants and resistance to herbivores: hypotheses, genetics, and variable responses in a diverse herbivore community. *Oecologia* **97:** 106-117.

Fox GA. 2001. Failure-time analysis. *Design and analysis of ecological experiments*: 253-289. Oxford University Pr.

Fuller R, McGhee K, Schrader M. 2007. Speciation in killifish and the role of salt tolerance. *Journal of Evolutionary Biology* **20**: 1962-1975.

Galloway LF, Fenster CB. 1999. The effect of nuclear and cytoplasmic genes on fitness and local adaptation in an annual legume, *Chamaecrista fasciculata. Evolution*. **53**: 1734-1743.

Gavrilets S, Li H, Vose MD. 2000. Patterns of parapatric speciation. *Evolution* **54**: 1126-1134.

Harrison S, Rajakaruna N. 2011. *Serpentine: the evolution and ecology of a model system*: Univ of California Pr.

Hatfield T, Schluter D. 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* **53**: 866-873.

Hendry A. 2004. Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evolutionary Ecology Research* **6**: 1219-1236.

Hendry A, Bolnick D, Berner D, Peichel C. 2009. Along the speciation continuum in sticklebacks. *Journal of Fish Biology*. **75**: 2000-2036.

Hendry A, Nosil P, Rieseberg L. 2007. The speed of ecological speciation. *Functional Ecology* **21**: 455.

Hereford J. 2009. A quantitative survey of local adaptation and fitness trade- offs. *The American Naturalist* **173**: 579-588.

Hoffmann AA, Merilä J. 1999. Heritable variation and evolution under favourable and unfavourable conditions. *Trends in Ecology & Evolution* **14**: 96-101.

Howard DJ. 1999. Conspecific sperm and pollen precedence and speciation. *Annual Review of Ecology and Systematics* **30**: 109–132.

Huberty AF, Denno RF. 2004. Plant water stress and its consequences for herbivorous insects: a new synthesis. *Ecology* **85**: 1383-1398.

Jain SK, Bradshaw AD. 1966. Evolutionary divergence among adjacent plant populations I. The evidence and its theoretical analysis. *Heredity* **21**: 407-441.

Kay KM. 2006. Reproductive isolation between two closely related hummingbird pollinated neotropical gingers. *Evolution* **60**: 538-552.

Keller I, Seehausen O. 2012. Thermal adaptation and ecological speciation. *Molecular Ecology* **21**: 782-799.

Kephart S, Theiss K. 2004. Pollinator- mediated isolation in sympatric milkweeds (*Asclepias*): do floral morphology and insect behavior influence species boundaries? *New Phytologist* **161**: 265-277.

Kirk H, Vrieling K, Van Der Meijden E, Klinkhamer P. 2010. Species by environment interactions affect pyrrolizidine alkaloid expression in *Senecio jacobaea*, *Senecio aquaticus*, and their hybrids. *Journal of Chemical Ecology* **36**: 378-387.

Kruckeberg A. 1986. An essay: the stimulus of unusual geologies for plant speciation. *Systematic Botany* **11**: 455-463.

Kruckeberg AR. 1951. Intraspecific variability in the response of certain native plant species to serpentine soil. *American Journal of Botany* **38**: 408-419.

Laland KN, Sterelny K, Odling-Smee J, Hoppitt W, Uller T. 2011. Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful? *Science* **334**: 1512-1516.

Langerhans RB, Gifford ME, Joseph EO. 2007. Ecological speciation in *Gambusia* fishes. *Evolution* **61**: 2056-2074.

Lowry DB. 2012. Ecotypes and the controversy over stages in the formation of new species. *Biological Journal of the Linnean Society* **106**: 241-257.

Lowry D, Modliszewski J, Wright K, Wu C, Willis J. 2008a. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B* **363**: 3009-3021.

Lowry DB, Rockwood RC, Willis JH. 2008b. Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution* **62**: 2196-2214.

Lowry DB, Willis J. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. 8:e100500.

Macnair M, Christie P. 1983. Reproductive isolation as a pleiotropic effect of copper tolerance in *Mimulus guttatus*. *Heredity* **50**: 295-302.

Martin NH, Bouck AC, Arnold ML. 2006. Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics* **172**: 2481-2489.

McBride CS, Singer MC. 2010. Field studies reveal strong postmating isolation between ecologically divergent butterfly populations. *PLoS Biology* **8**: e1000529.

McKay JK, Richards JH, Mitchell-Olds T. 2003. Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Molecular Ecology*. **5**: 1137-1151.

McNeilly, T., and J. Antonovics. 1968. Evolution in closely adjacent plant populations IV. Barriers to gene flow. *Heredity* **23**: 205-218.

Miglia KJ, Mcarthur ED, Moore WS, Wang H, Graham JH, Freeman D. 2005. Nine- year reciprocal transplant experiment in the gardens of the basin and mountain big sagebrush (*Artemisia tridentata*: Asteraceae) hybrid zone of Salt Creek Canyon: the importance of multiple- year tracking of fitness. *Biological Journal of the Linnean Society* **86**: 213-225.

Moyle LC, Graham EB. 2005. Genetics of hybrid incompatibility between *Lycopersicon esculentum* and *L. hirsutum*. *Genetics* **169**: 355-373.

Nagy E, Rice K. 1997. Local adaptation in two subspecies of an annual plant: implications for migration and gene flow. *Evolution* **51**: 1079-1089.

Nosil P. 2004. Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**: 1521-1528.

Nosil P. 2007. Divergent host plant adaptation and reproductive isolation between ecotypes of *Timema cristinae* walking sticks. *The American Naturalist* **169**: 151-162.

Nosil P, Egan SP, Funk DJ. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: ‘isolation by adaptation’ and multiple roles for divergent selection. *Evolution,* **63**: 316-336.

Nosil P, Harmon L, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* **24**: 145-156.

Nosil P, Vines T, Funk D. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**: 705-719.

O'Dell RE, Rajakaruna N. 2011. Intraspecific variation, adaptation, and evolution. *Serpentine: The Evolution and Ecology of a Model System*.

Ornduff R. 1964. Evolutionary pathways of the *Senecio lautus* alliance in New Zealand and Australia. *Evolution* **18**: 349-360.

R Development Core Team. 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Viena, Austria. ISBN 3-900051-07-0, URL http://www/R-project.org.

Radford I, Cousens R, Michael P. 2004. Morphological and genetic variation in the *Senecio lautus* complex: are variants worthy of taxonomic recognition? *Australian Systematic Botany* **17**: 29-48.

Ramsey J, Bradshaw Jr H, Schemske D. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**: 1520-1534.

Roda F, Ambrose L, Walter GM, Liu H, Schaul A, Lowe A, Pelser PB, Prentis P, Rieseberg LH, Ortiz-Barrientos D. 2013a. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Molecular Ecology* **22**: 2941-2952.

Roda F, Liu H, Wilkinson M, Walter G, James M, Bernal D, Melo M, Lowe A, Rieseberg L, Prentis P, Ortiz-Barrientos D. 2013b. Convergence and divergence during the adaptation to similar invironments by an Australian groundsel. *Evolution* **67**: 2515-2529.

Rundle HD, Nosil P. 2005. Ecological speciation. *Ecology Letters* **8**: 336-352.

Rundle HD, Whitlock MC. 2001. A genetic interpretation of ecologically dependent isolation. *Evolution* **55**: 198-201.

Sambatti J, Rice KJ. 2007. Functional ecology of ecotypic differentiation in the Californian serpentine sunflower (*Helianthus exilis*). *New Phytologist* **175**: 107-119.

Sambatti J, Strasburg JL, Ortiz-Barrientos D, Baack EJ, Rieseberg LH. 2012. Reconciling extremly strong barriers with high levels of gene exchange in annual sunflowers. Evolution **66**: 1459-1473.

Schemske DW. 2010. Adaptation and the origin of species. *American Naturalist* **176**: S4-S25.

Schluter D 1998. Ecological causes of speciation. *Endless forms: species and speciation.* New York: Oxford Univ. Press.

Schluter D. 2000. *Ecology of adaptive radiation*: Oxford University Press.

Schluter D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* **16**: 372-380.

Silvertown J, Servaes C, Biss P, Macleod D. 2005. Reinforcement of reproductive isolation between adjacent populations in the Park Grass Experiment. *Heredity* **95**: 198-205.

Sork VL, Stowe KA, Hochwender C. 1993. Evidence for local adaptation in closely adjacent subpopulations of Northern Red Oak (*Quercus rubra L.*) expressed as resistance to leaf herbivores. *The American Naturalist* **142**: 928-936.

Stebbins GL. 1952. Aridity as a stimulus to plant evolution. *American Naturalist* **6**: 33-44.

Thibert- Plante X, Hendry A. 2009. Five questions on ecological speciation addressed with individual- based simulations. *Journal of Evolutionary Biology* **22**: 109-123.

Thibert-Plante X, Hendry AP. 2011. The consequences of phenotypic plasticity for ecological speciation. *Journal of Evolutionary Biology* **24**: 326-342.

Thompson I. 2005. Taxonomic studies of Australian *Senecio* (Asteraceae): 5. The S. lautus/S. lautus complex. *Muelleria* **21**: 23-76.

Vekemans X, Lefèbvre C. 1997. On the evolution of heavy- metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers. *Journal of Evolutionary Biology* **10**: 175-191.

Via S, Bouck A, Skillman S. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* **54**: 1626-1637.

Westberg E, Poppendieck H-H, Kadereit JW. 2010. Ecological differentiation and reproductive isolation of two closely related sympatric species of *Oenanthe* (Apiaceae). *Biological Journal of the Linnean Society* **101**: 526-535.

White EM. 2008. *Indirect interactions between alien and native Senecio species as mediated by insects.* PhD by Publication, Queensland University of Technology.

Wu L, Bradshaw A, Thurman D. 1975. The potential for evoltution of heavy metal tolerance in plants. III. The rapid evolution of copper tolerance in *Agrostis stolonifera. Heredity* **34**: 165-187.

Tables

Table 2.1. Intrinsic RI in *Senecio lautus* measured in the glasshouse.

-patry	Locality	[a]Cross	[b]N	[c]RI
Para	**Hat Head**	HxD	25	-0.053
		DxH	25	0.241
	Lennox Head	HxD	54	-0.145
		DxH	60	0.098
	Cabarita Beach	HxD	25	-0.089
		DxH	25	0.015
Allo	**Cabarita Beach/ Byron bay**	DxH	27	0.056
		HxD	27	-0.221
	Lamington NP/Port Macquarie	TxH	20	-0.230
		HxT	20	-0.166

[a] D, H, and T refer to Dune, Headland, and Tableland population.

[b] N is the number of families that were crossed.

[c] RI is the strength of postmating prezygotic RI measured as the relative fecundity of hybrids (F1 seed set) over that of parental types (parental seed set) in three parapatric and two allopatric populations.

Table 2.2. Proportion of germination for each experimental cross in reciprocal transplant experiments in Cabarita Beach.

Habitat	[a]Cross type	N	[b]G1	[c]G2
Dune	**D**	275	30.59	34.18
	F1-D	159	52.12	54.08
	F1-H	159	37.15	40.25
	H	269	29.38	30.48
Headland	**D**	271	10.75	12.54
	F1-D	151	21.89	22.51
	F1-H	153	18.98	23.52
	H	265	14.78	18.11

[a]Cross types are Dune (D) and Headland (H) parental types and reciprocal F1 hybrids where letters denote the identity of the mother (F1-H and F1-D)

[b]G1 is the proportion of seeds that germinated during the first days of the experiment.

[c]G2 refers to a second bout of germination (day 135 in the Headland environment, and day 145 in the Dune environment).

Table 2.3. Individuals with flowers and their average number of flowers for each experimental cross in reciprocal transplant experiments in Cabarita Beach.

Habitat	[a]Cross type	Seeds sowed	Alive at first day of flowering	Mean number of flowers per individual	S.E.
Dune	**D**	275	28	0.0641	0.0447
	F1-D	159	15	0.0074	0.0073
	F1-H	159	13	0	0
	H	269	6	0	0
Headland	**D**	271	8	2.6475	2.2360
	F1-D	151	8	1.1260	0.5569
	F1-H	153	12	2.0125	1.1435
	H	265	17	1.0425	0.5325

[a]Cross types are Dune (D) and Headland (H) parental types and reciprocal F1 hybrids where letters denote the identity of the mother (F1-H and F1-D).

Table 2.4. Strength of RI barriers between Cabarita Beach populations.

[a]**Reproductive isolating barrier**	**Habitat**	[c]**Cross**	**Strength**
Flowering asynchrony	Dune	-	0.3939
	Headland	-	0.48
Immigrant inviability	Dune	-	0.78
	Headland	-	0.54
[b]**Extrinsic postzygotic**	Dune	F1-D	0.41
		F1-H	0.32
		F1	0.37
	Headland	F1-D	0.34
		F1-H	0.06
		F1	0.19
Hybrid F1 seed set	Glasshouse	F1-D	0.015
		F1-H	-0.089
		F1	-0.038
Intrinsic hybrid viability	Glasshouse	F1-D	0
		F1-H	0
		F1	0
Average prezygotic	Dune		0.59
Average postzygotic			0.11
Cumulative			0.877
Average prezygotic	Headland		0.55
Average postzygotic			0.038
Cumulative			0.76

[a]Estimates were calculated following Lowry *et al.*, 2008a approach.

[b]Extrinsic postzygotic isolation was calculated from either the total number of seeds or from individuals that germinated in the reciprocal transplant experiment, respectively.

[c]Cross types are Dune (D) and Headland (H) parental types and reciprocal F1 hybrids where letters denote the identity of the mother (F1-H and F1-D).

Table 2.5. Magnitude of local adaptation (local adaptation index) at each habitat for each experimental cross in reciprocal transplant experiments in Cabarita Beach.

		Magnitude of Local Adaptation		
Habitat	**Local Vs. non-local [a]cross type**	**Viability**	**Fecundity**	**Composite**
Dune	**D Vs. H**	0.96	2.56	3.05
	D Vs. F1-H	0.61	2.56	3.05
	D Vs. F1-D	0.70	1.12	2.10
Headland	**H Vs. Dune**	0.55	-1.08	-0.35
	H Vs. F1-D	0.40	0.07	0.44
	H Vs. F1-H	0.23	-0.40	-0.19

[a]Cross types are Dune (D) and Headland (H) parental types and reciprocal F1 hybrids where letters denote the identity of the mother (F1-H and F1-D).

Figures

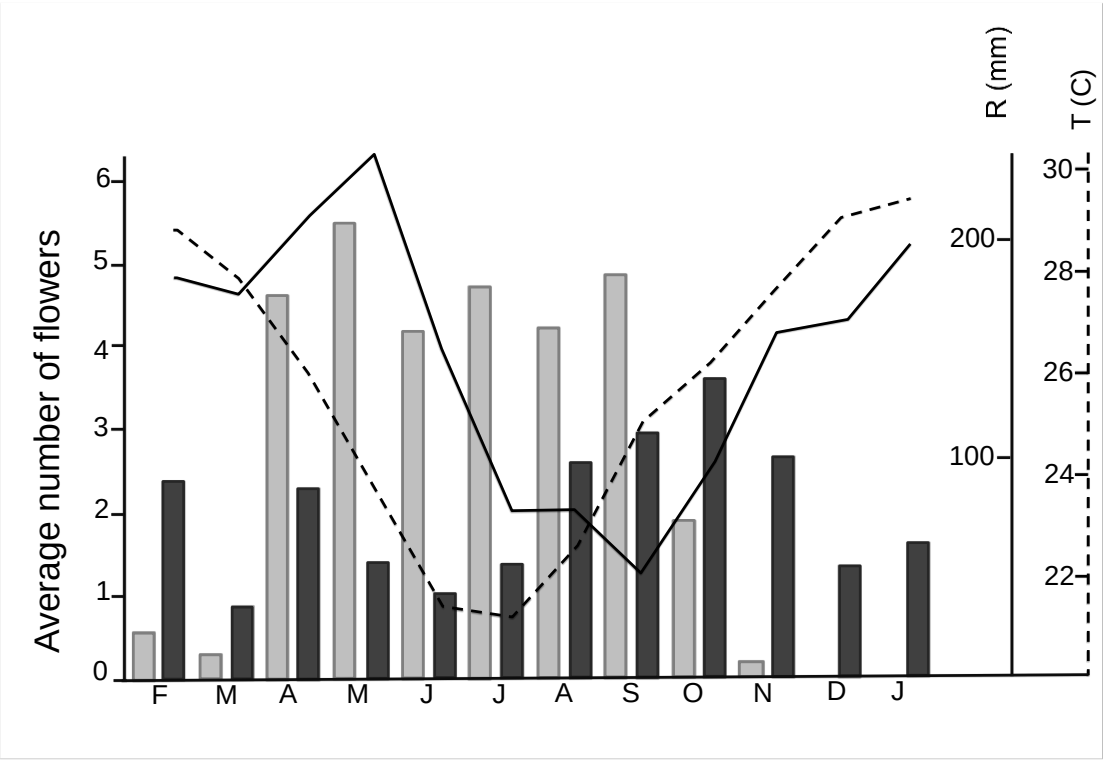

Fig. 2.1. Average number of flowers per individual from the natural populations in the sand dunes (light grey bars) and rocky headlands (dark grey bars) through 12 consecutive months (February 2011 to January 2012). Monthly average rainfall (mm) (dark line) and temperature in C (dashed line) are averages over the past 10 years.

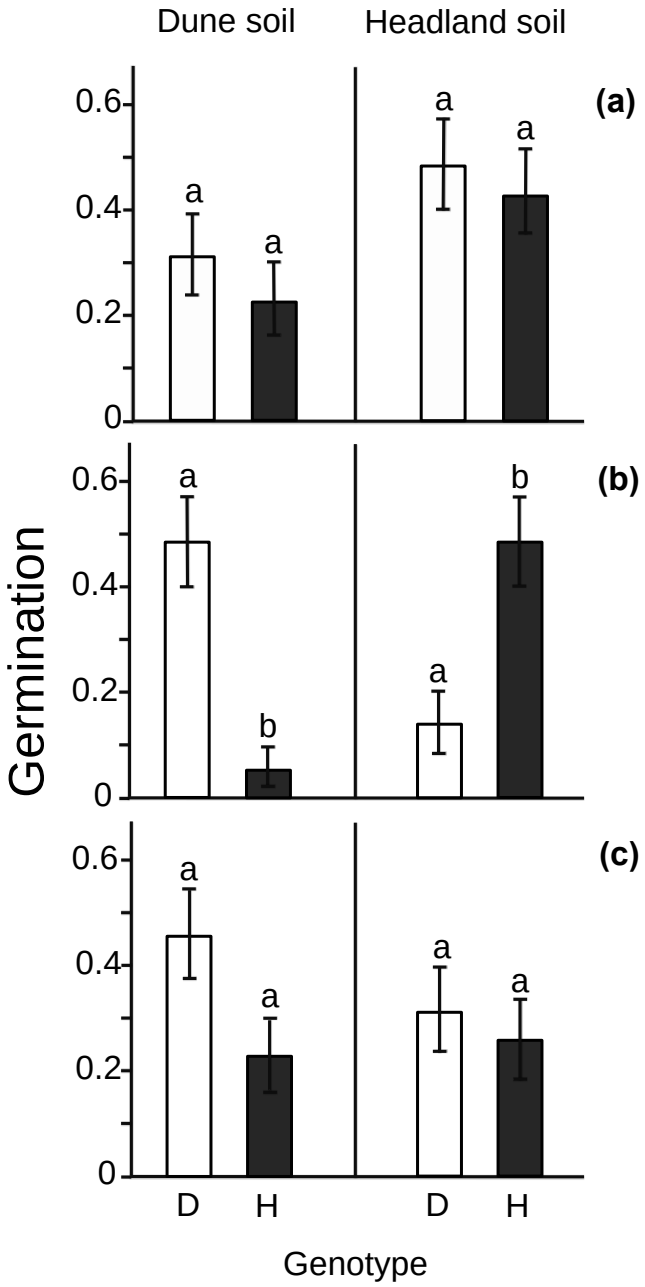

Fig. 2.2. Proportion of Dune and Headland seeds that germinated in Dune soil (left panels) and Headland soil (right panels) collected from three different localities: (a) Cabarita Beach, (b) Lennox Head and (c) Stradbroke Island. Bars show standard errors with letters letters denoting significant differences. These experiments were conducted under controlled environmental conditions in the glasshouse

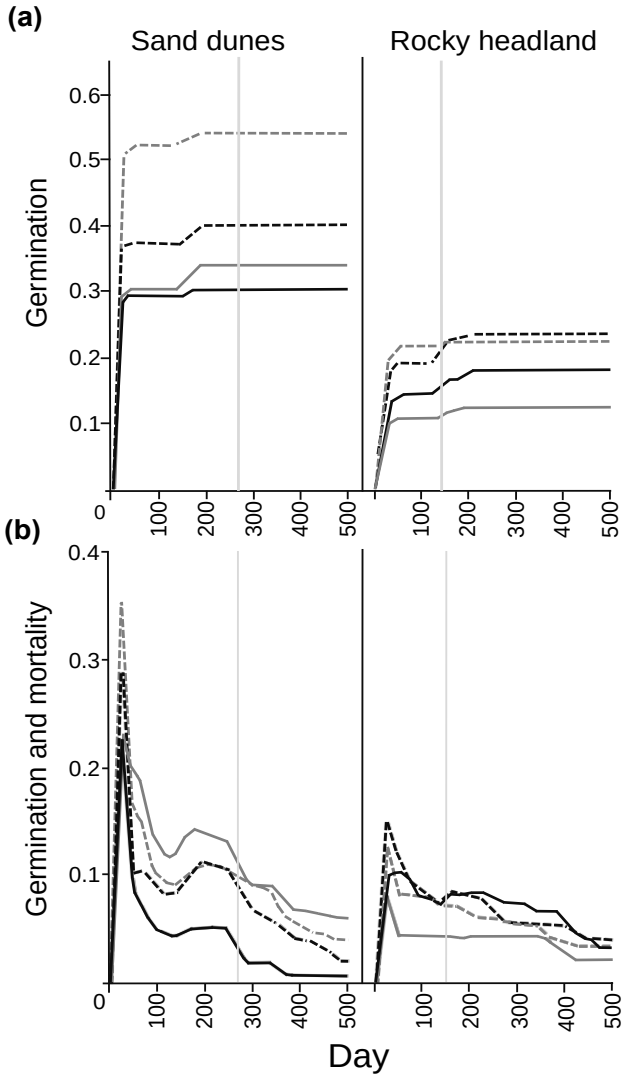

Fig. 2.3. Germination and mortality for parental and hybrid types in field transplant experiments. (a) Proportion of seeds that germinated in experimental block in the sand dunes and rocky headland of Cabarita Beach. (b) Combined effects of germination and mortality on the total number of survivors relative to the total number of seeds planted. Vertical lines indicate the day at which the first flower appeared in each environment (day 275 in the sand dunes and day 149 in the rocky headland). Black and grey lines correspond to the Headland and Dune crosses respectively and the dashed black and gray lines to F1-H (Headland mother) and F1-D (Dune mother) crosses respectively.

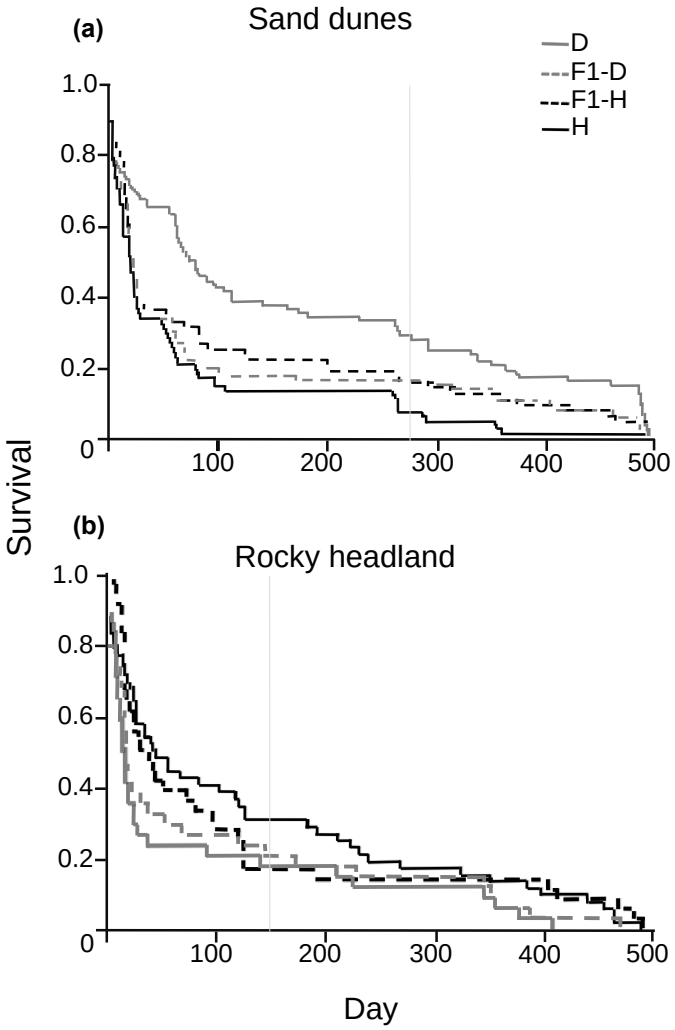

Fig. 2.4. Survival of parental and hybrid types in the (a) sand dunes and (b) rocky headland of Cabarita Beach. Vertical lines indicate the day at which the first flower appeared at each environment (day 275 in the sand dunes and day 149 in the rocky headland).

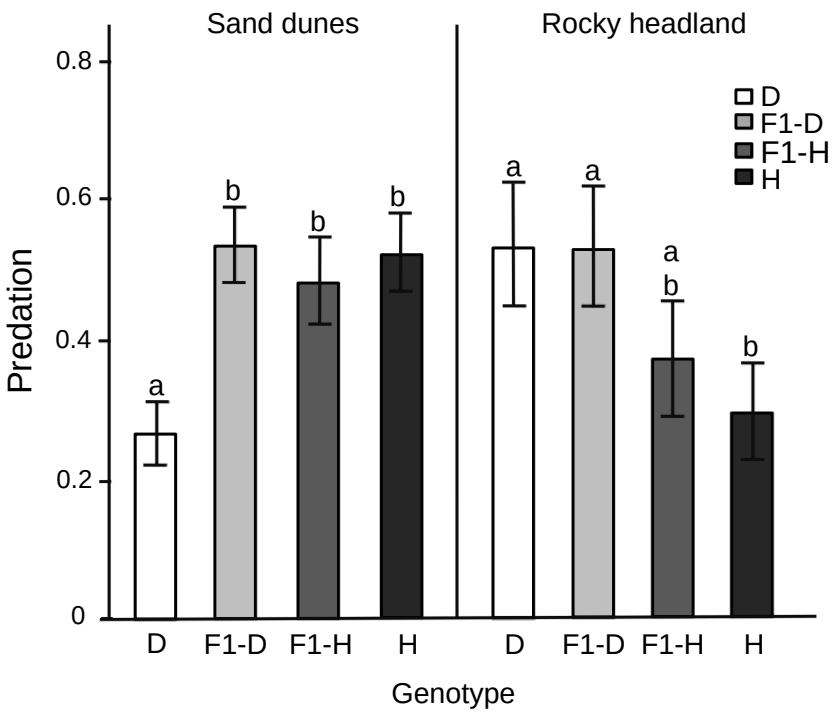

Fig. 2.5. Mortality due to predation in in the sand dunes and rocky headland of Cabarita Beach. Bars show standard errors for binomial probabilities, with letters denoting significant differences. Bars ranging in color from white to dark grey represent the Dune, F1-D, F1-H, and Headland cross types, correspondingly.

CHAPTER III

LACK OF GENE FLOW IN PARAPATRIC *SENECIO LAUTUS*

ABSTRACT

Ever since Darwin, the role of natural selection on the origin of new species has been ardently debated and studied. Although recent empirical and theoretical work suggests that natural selection is a key driver of speciation, we still remain largely ignorant as to the consequences of speciation on overall patterns of genome divergence and gene flow between populations. Notably, it remains unclear weather natural selection can reduce gene flow between parapatric populations adapting to contrasting environments. Here we evaluate the consequences of natural selection on gene flow between six parapatric population pairs and four allopatric coastal populations of the *Senecio lautus* complex. We estimate levels of gene flow with coalescent multilocus approaches and examine data from population genomic scans across populations of the complex. Unexpectedly, we found that levels of gene flow are drastically reduced between parapatric populations, despite higher heterogeneous genomic divergence in parapatry than in allopatry. We suggest that natural selection drove sudden reductions of gene flow between parapatric populations, thus converting them in to ecologically allopatric populations.

Key words: Parapatry, gene flow, migration, reproductive isolation, polymorphisms, neutral loci.

INTRODUCTION

Ecological speciation is the process in which reproductive isolation (RI) evolves between populations adapting to contrasting environments (Schluter, 2000; Schluter, 2001; Rundle & Nosil, 2005). If populations are geographically isolated (allopatry), alleles underlying adaptive traits can easily fix. But when populations are in contact (sympatry and parapatry), interspecific recombination tends to dissolve adaptive allelic arrangements making the early stages of speciation with gene flow unlikely (Felsenstein, 1981; Seehausen *et al.*, 2014). A growing body of well-documented cases in both animals and plants [apple maggot flies *Rhagoletis pomonella* (Filchak *et al.*, 2000), benthic-limnethic species of the threespine stickleback (Hatfield & Schluter, 1999), *Timmema* walking sticks (Nosil *et al.*, 2008), *Littorina* marine snails (Butlin *et al.*, 2008), ecotypes of *Eucalyptus globulus* (Foster *et al.*, 2007), and several plant populations and species inhabiting serpentine soils (Harrison & Rajakaruna, 2011)] plus several theoretical models (e.g., Gavrilets *et al.*, 2000; Bolnick & Fitzpatrick, 2007; Thibert- Plante & Hendry, 2009), suggest that ecological speciation in the face of gene flow could be common. However, it remains unclear the extent to which natural selection can reduce gene flow between populations, and thus whether the evolution of extrinsic reproductive isolation (see below) alone can drive speciation (Seehausen *et al.*, 2014).

Geographic modes of population divergence govern the way in which genetic differentiation accumulates, and largely determine the reproductive isolating barriers that are likely to arise at different stages of the speciation process (Feder *et al.*, 2013; Seehausen *et al.*, 2014). When ecological divergence occurs in parapatry, natural selection favours genetic changes at loci responsible for relevant ecological traits while the rest of the genome is shared between populations through gene flow (Schluter & Conte, 2009). Adaptive changes are usually specific to the local environment but maladaptive in the alternative one, resulting in the evolution of reproductive isolating barriers dependent on the environment (extrinsic), such as immigrant inviability (selection against immigrant individuals; Nosil *et al.*, 2005) and extrinsic postzygotic isolation (selection against non-adapted hybrids; Rundle & Whitlock, 2001). Intrinsic reproductive isolation, such as hybrid sterility or inviability, is expected to evolve later in the process, as these barriers usually evolve under minimal levels of gene flow between populations. Distinctively, when ecological divergence occurs in allopatry, both natural selection and genetic drift can drive the accumulation of genetic incompatibilities between populations and thus cause intrinsic hybrid dysfunctions during the early stages of speciation (Turelli *et al.*, 2001; Coyne & Orr, 2004). In contrast to the parapatric case, both intrinsic and extrinsic RI can be expected to evolve early during the allopatric process (Seehausen *et al.*, 2014).

Studies often find it challenging to distinguish between speciation initiated by divergent natural selection in parapatry or sympatry and speciation initiated in allopatry followed by secondary contact (Coyne & Orr, 2004). One possible explanation for this difficulty is that in advanced stages of ecological speciation, isolated populations by extrinsic reproductive barriers can accumulate genetic differentiation due to genetic drift (Thibert Plante & Hendry, 2010). This process could leave similar genetic signatures to those found in populations that diverged in allopatry but exchanged genes after secondary contact (Hendry, 2009). Similarly, many of the differences accumulated in allopatry can be lost due to gene flow after secondary contact, thus leading to patterns of heterogeneous genomic divergence similar to those predicted under parapatric and sympatric speciation (Dieckmann & Doebeli, 1999; Gavrilets, 2003; Bolnick & Fitzpatrick, 2007; Feder *et al.*, 2013). In addition, in most cases it is difficult to evaluate whether populations are under migration-selection balance or selection has tipped the system to reduced levels of gene flow facilitating the evolution of reproductive isolation (e.g., intrinsic; Gavrilets, 2004). As a consequence, in most systems it remains difficult to distinguish "allopatry-first" versus "allopatry-second (effective allopatry resulting from the evolution of strong extrinsic reproductive isolation in the face of gene flow, and see Fig. 1.1)," and thus it is challenging to discover how reproductive isolation and reductions in gene flow evolve through the speciation continuum (but see Powell *et al.*, 2013; Seehausen *et al.*, 2014 for an example and a review, respectively).

Systems where ecotypes or species evolve in parallel provide an appropriate arena to study the effect of natural selection and gene flow on different levels of population divergence and speciation. The repeated and independent evolution of traits across multiple populations strongly implicates natural selection as the driver for trait differentiation between populations, a process that would be unlikely under the vagaries of genetic drift (Schluter & Nagel, 1995; Ostevik *et al.*, 2012). If adaptive traits also confer ecotypes with reproductive isolation, then population may speciate in parallel. Systems evolving through parallel ecological speciation, should ideally meet three criteria: 1) phylogenetic independence of each diverging pair, 2) reproductive isolation between populations at contrasting environments and 3) reproductive compatibility between populations inhabiting the same environments (Schluter & Nagel, 1995; Ostevik *et al.*, 2012). Good candidate systems for parallel ecological speciation are freshwater sticklebacks (Rundle *et al.*, 2000; Hohenlohe *et al.*, 2010), freshwater salmons (Perrier *et al.*, 2013), *Timema* walking sticks (Nosil *et al.*, 2002), dwarf populations of *Eucalyptus globulus* (Foster *et al.*, 2007), and serpentine populations of *Cerastium alpinum* (Berglund *et al.*, 2004), with the evidence being more common and stronger in animals than in plants (Ostevik *et al.*, 2012).

Here, we use the *Senecio lautus* species complex to explore how natural selection has affected patterns of gene flow in multiple morphologically differentiated parapatric populations that have evolved repeatedly and independently along the coast of Australia. We use neutral DNA sequences to directly measure levels of gene flow between populations and examine of patterns of shared polymorphism at a genomic scale. We discuss our results in terms of the main alternative speciation continuums (Fig. 1.1) and conclude that natural selection is likely to have reduced gene flow to nil levels across the system.

S*enecio lautus* is a young groundsel ecotypic and species complex comprising populations and species adapted to a wide variety of environments in Australia (Ali, 1966; Ali, 1968; Ali, 1969; Radford *et al.*, 2004; Thompson, 2005). Within the complex, populations adapted to parapatric sand dunes (Dune, or D) and rocky headland (Headland, or H) habitats show genetically based phenotypic differences, where Dune individuals are tall, erect and poorly branched, and Headlands are short, prostrated and heavily branched (Thompson, 2005). The two ecotypes are largely inter-fertile in glasshouse conditions (Ali, 1964; Ali, 1968), but see Chapter IV for recent discoveries, are evolving under divergent natural selection (Melo *et al.*, 2014), and have arisen repeated times along the eastern and southern coast of Australia (Roda *et al.*, 2013). Although much of its biology and geography, as well as recent published analyses of heterogeneous genomic divergence in these coastal forms suggests that they have evolved in the presence of gene flow, we lack direct estimates of gene exchange between parapatric populations, thus limiting our understanding as to how selection shapes divergence during the early stages of speciation.

MATERIALS AND METHODS

Study populations and sample preparation

Six parapatric Dune and Headland populations, two Inland and two Alpine populations were selected for this study (for geographic coordinates see Table 3.1 and map in Fig. 3.1a). *Senecio madagascariensis*, a closely related species to *S. lautus* from Africa (Radford & Cousens, 2000; Thompson, 2005) was included as an outgroup in some analyses (see below), including HKA neutrality and phylogenetic analyses (Roda *et al.*, 2013). We sampled 12 individuals (24 chromosomes) per each Dune and Headland, and 11 for Alpine and Inland population. Leaf tissue for DNA extraction was stored at -80ºC in the Ortiz-Barrientos Laboratory at The University of Queensland, Australia. A modified CTAB protocol was used for DNA extractions, leading to DNA samples with final concentration of 30ng/μL (Roda *et al.*, 2013).

Library construction and sequencing

We prepared DNA libraries for each of the 192 individuals for targeted re-sequencing in the Fluidigm Access Array system (Moonsamy *et al.*, 2011). This is an efficient 4-primer PCR process using fluidics, where every individual is 'simultaneously tagged' with a specific nucleotide sequence as PCR products accumulate. This enables samples to be pooled for next-generation sequencing and tracked subsequently for bioinformatic and population genetic analyses. We prepared libraries for 96 genomic regions, including genes previously described in Roda *et al*. 2013b, and the internal transcribed spacer (ITS) between the 5.8s and 18s ribosomal RNA genes. Primers were designed in BatchPrimer3. Each target primer also included a common sequence adapter that matched a second set of primers that carry the barcode sequence and the technology specific adapter for next-generation sequencing (e.g., Roche 454 adapters in our case). This double primer system facilitates pair-ended sequencing as well as individual tracking across all sequenced loci in the Access Array system (Moonsamy *et al.*, 2011). The access array is a miniaturised PCR machine that allows construction of 48 DNA libraries for at least 48 loci in simultaneous (multiplexing can lead to greater number of loci incorporated into the library). Products can then be sequenced in a next generation sequencing platform. Before library preparation in the Access Array, we validated amplification of 96 loci, individually, and by multiplexing two at a time. We followed a shortened Access Array PCR amplification protocol (see Appendix 3 for details). After validation, we prepared a total of 192 libraries (four access array PCR chips) each containing barcoded pools from PCR products of up to 96 loci following the specifications of the Access Array manual (Moonsamy *et al.*, 2011). All libraries were pooled in equimolar quantities and sent for Next Generation sequencing at the Beijing Genomics Institute (BGI). Pools were further cleaned with an Agencourt amPure purification kit to remove short sequencing products and primer dimers. Cleaned pools were sequenced using the emPCR (Lib-A) for bi-directional sequencing kit on the Roche GS FLXTitanium platform.

Read processing

After trimming barcode complexes with TagCleaner (Schmieder *et al.*, 2010), reads that had more than 40% low quality (<q20) pb, had Ns higher than 2% or were shorter than 50 bp; were eliminated from the subsequent analyses. PRGmatic (Hird *et al.*, 2011) was used for read alignment, where 100bp was the least overlap required (Hird *et al.*, 2011). CAP3 (Huang & Madan, 1999) was used to cluster reads with a 99% and 90% similarity threshold within individuals and across all individuals respectively and used to construct a reference genome. Reads were then aligned to this genome reference using BWA (Li & Durbin, 2009). Alignments were sorted by

SAM tools (Li *et al.*, 2009), allowing SNP frequency to be calculated; whilst VarScan (Koboldt *et al.*, 2009) detected insertions and deletions and called the SNPs. PRGmatic constructed a haplotype for each locus and individual, based on the threshold for assigning a true SNP (minimum coverage of 3). Haplotypes were divided into a separate file per locus (containing all individuals with the corresponding haplotype sequence). A contig, or 'locus' was only considered for further analyses if it had haplotypes assigned in 6 or more individuals per populations. Loci from PRGmatic were examined in relation to the expected amplicons using BLAT (Kent, 2002). If an amplicon had more than one locus mapping to it, it was subjected to further investigation. If the reads within the locus that mapped to a single amplicon were found (upon visual inspection) to be either 1) highly variable (i.e., many single base pair differences between the reads) 2) slightly variable, but each contig contained almost all of the same individuals, or 3) a segment of the reads across population was quite similar, the other region being highly variable, they were excluded from the final data set. These instances possibly suggest pseudo-genes, duplicated genes or a recombination event respectively. High quality loci were concatenated and aligned in MUSCLE (Edgar, 2004). 26 loci were of high quality and were present in all the populations in this study.

DNA sequence polymorphism analyses

We conducted a basic polymorphism analyses using DNAsp (Librado & Rozas, 2009). Number of haplotypes, haplotype diversity, and nucleotide diversity as the average number of nucleotide differences per site between sequences pairs (π; Nei, 1987) and per base pair (Theta; Watterson, 1975) were estimated excluding Indels (Table S3.1). Deviations from neutrality (Kimura, 1985) for each gene at each population were tested using HKA (Hudson *et al.*, 1987), Tajima's D (Tajima, 1989) and Li and Fu's (Fu & Li, 1993) tests. Studies on the different tests statistical power suggest HKA is more powerful than other tests based on population genetic data (Zhai *et al.*, 2009). However, the lack of sequences for *S. madagascariensis* prevented us from performing this test in some of the loci (See Table S3.1). Loci where we did not perform the HKA test were considered to reject the neutral model if the Fu and Li's tests as well as Tajima's D suggested it.

Phylogenetic analysis population differentiation for neutral markers

A Bayesian phylogenetic analysis was performed using the program *Beast, an extension of Beast 1.7.5 for species tree estimation (Heled & Drummond, 2010). Analysis was performed with a chain 300,000,000 long, using a strict molecular clock, which assumes a global clock rate with no variation among lineages in a tree. ITS was used as the reference locus –with mutation rate for herbaceous plants of $4.13*10^{-9}$ subs/site/year (Kay *et al.*, 2006)– from which the rate of the other genes was estimated. According to Bayesian Information Criterion values found using jModeltest

(Posada & Crandall, 1998), the best substitution model for 17 out of the 26 genes was the HKY model (Hasegawa *et al.*, 1985). We used a *Yules* species process for species tree estimation (this assumes that lineages split at a constant rate). We used Mantel tests (Oksanen *et al.*, 2013) in R (R Core Team, 2012) to evaluate correlation between divergence times and geographic distance between all populations.

To investigate population differentiation, three different methods were applied using common neutral genes common across populations. First, an AMOVA was used to partition molecular variance into different hierarchical levels (Excoffier *et al.*, 1992) using the program Arlequin ver. 3.1 (Excoffier *et al.*, 2005). Second, we tested for a correlation between geographic distance with F_{ST} estimates for all possible population pairs performing a permutation approach for the Mantel test as implemented in the R package *vegan* (Oksanen *et al.*, 2013). Finally, the potential number of genetic clusters (K) was inferred using STRUCTURE v2.3.4 (Pritchard *et al.*, 2000). We explored population structure and levels of admixture across the species complex at two levels: *i)* including all populations and *ii)* by pairs (6 parapatric and 2 allopatric). STRUCTURE was run using an admixture model to allow a fraction of an individuals genome to have mixed ancestry, and the correlated allele frequency model to account for correlations between linked markers (Falush *et al.*, 2003). Parameters used for *i)* K=1-16, 20 iterations per K, burnin 100,000, MCMC 100,000; and for *ii)* K=1-9, 20 iterations per K, burnin 100,000, MCMC 100,000, as suggested in Kimberly *et al.* 2012. To choose the most likely K value, both methods by Pritchard *et al*. (2000) and Evanno *et al*. (2005) were examined. Because both methods tend to overestimate K, and high K values did not add any additional clustering information compared to smaller Ks, the smallest K that captured the major structure in the data was chosen, ensuring that all summary statistics converged.

Gene Flow

Gene flow estimates were derived from neutral and non-recombining DNA sequences. Inter-loci recombination was assessed based on the four-gamete test (Hudson & Kaplan, 1985) as implemented in DNAsp (Librado & Rozas, 2009). For loci where recombination events were detected, we trimmed off the shortest fragment and included the longest into the analysis, as sequences including recombination events could lead to a false signature of gene flow (Hey & Nielsen, 2004). Estimates of gene flow were assessed using IMa2 (Hey, 2010) making pair wise comparisons between the six parapatric pairs and two allopatric comparisons (see Fig. 3.1a). Gene flow between proximate populations inhabiting the same kind of habitat was also estimated (between Dunes and between Headlands). We used 100,000 burn-in steps for a minimum of 100,000 genealogies for parameter estimation. Runs involved 150 to 180 independent chains,

effective sample size (ESS) among all parameters ranged from several hundred to >10,000. Upper bounds of the prior distributions for each parameter were set based on the results of a preliminary run. Mutation rates for each of the genes in the analysis (Table 3.2) were obtained using the relationship D=2*u*T, where D is the average number of substitutions separating *Senecio lautus* complex of *S. madagascariensis*, and T is the age of the *S. lautus* clade (~ 150, 000 years; Roda *et al.*, 2013a). However, gene flow estimates did not change if we ignored mutation rates (data not shown). We ran three independent IMa2 analyses varying seed number to ensure parameter estimation was consistent. In addition to the estimates of gene flow between parapatric pairs and some selected allopatric populations, we also estimated levels of gene flow between the closest populations inhabiting the same environment (e.g. comparisons between Dune or between Headland populations).

RESULTS

Patterns of DNA sequence variability

We studied DNA sequence polymorphism in 26 loci from 16 populations of *Senecio lautus* distributed across the eastern and southern coasts of Australia (Fig. 3.1a; Table 3.1). The average length for all genes after removing indels was 382.88 +/- 7.37 bp, with the shortest locus 278 bp long and the longest 437 bp. The average number of segregating sites per locus was 4.72 +/- 0.22 (see Table S3.2 for full details). Overall, 1199, 1091, 1125, 1021 and 1018 are amongst loci with the highest variability in terms of π and Theta (Table 3.2). Contrastingly, loci 1116, 1176, 1204 and 1212 were amongst the ones with lowest DNA sequence variability. Dune and Headland ecotypes had similar average values of π and Theta, but Alpine and Inland ecotypes had higher values of Theta (Fig. 3.2). All genes had on average negative values for Tajima's D (Table 3.2). Amongst 416 estimates of Tajima's D (Table S3.2) for each gene (26) and each population (16), only 20 cases had positive values for it, although they were not significantly different from zero (Table S3.2). In particular, population H01 contained seven loci exhibiting positive values. Table S3.2 reports the length of each of these genes (removing indels), the number of sequences available at each population, and polymorphism statistics. The last are summarized by loci and across all populations in Table 3.2.

Neutral tests

We performed the HKA (Hudson *et al.*, 1987) test for 14 of the 26 genes in the study, for which neutrality was assessed only based on this test result. For the remaining 12 loci sequences with out sequence for the outgroup, we assessed neutral evolution based on the results of Fu & Li's tests D*

and F* (Fu & Li, 1993) and Tajima's D (Tajima, 1989). In the few cases we found discordances between the three tests, loci were considered neutral if either the two Fu and Li's test supported neutrality or the Tajima's D alone did (see Table S3.1 for details on the results for each of the neutrality tests). A total of 13 loci were considered neutral at all populations (Table 3.3 and S3.1). The number of neutral loci for pairwise comparisons ranged from 20 to 24 (Table 3.3). Those genes showing evidence of selection in some populations were previously found to be involved in genetic differentiation across multiple parapatric pairs (1084, 1085, 1090, 1091, and 1098), flowering and reproduction (e.g., 1004, 1018, and 1021), and plant architecture (1125 and 1126).

Phylogenetic relations and genetic structure

Phylogenetic analyses using 13 neutral loci common to all populations clustered *Senecio lautus* populations according to geographic distribution (Table 3.3 and Fig. 3.1). As shown in previous work (Roda *et al.*, 2013a,b) we confirmed the existence of two main sister clades in the complex (a southern and an eastern clade) and each coastal pair formed an independent monophyletic group (Fig. 3.1). Both clades displayed phylogenetic substructure that matched geography and not habitat or morphology. Populations from South Australia as well as populations from southern Victoria and Tasmania formed monophyletic subclades. Inland populations were sorted into the northern and southern clade, but Alpine populations were grouped together in the southern clade (Fig. 3.1b).

STRUCTURE (Pritchard *et al.*, 2000) analysis revealed that the most likely number of genetic clusters in the entire system was two, corresponding to main phylogenetic clades (Fig. 3.1c, but note A03). To explore differentiation between populations we ran the same STRUCTURE analyses in each coastal pair, and for two additional allopatric comparisons. We found that most population pairs were not differentiated, with H01 and D01 displaying the strongest subdivision (Fig. 3.1). Despite strong phylogenetic signal and population structure, Analyses of Molecular Variance (AMOVA) showed that the highest degree of variation came from differences found amongst individuals within populations (89.45%), and very little amongst clades (4%), or amongst populations within clades (6.56%).

We found a positive and strong correlation between Fst and geographic distance (Mantel test: r = 0.5463, p = 0.001). Fixation indices, Fsc and Fst, were significant for all genes except for locus 1176 (Table S3.3, Table S3.4 and for AMOVA results per gene). Estimated divergence times using the average mutation rate for herbaceous plants (Kay *et al.*, 2006) indicated that *Senecio lautus* is a young species complex that originated less than 500,000 years ago. The youngest divergence between a parapatric Dune and Headland pair was estimated to be 75,000 years, and the oldest 250,000 years (Fig. 3.1b). Genetic divergence correlated positively and strongly with geographic

distance (Mantel test: $r = 0.7806$, $p = 0.001$), suggesting that maximum genetic divergence plateaus around 500 km.

Patterns of gene flow between parapatric populations

We used coalescent models of isolation with migration (IMa2; Hey & Nielsen, 2004; Hey & Nielsen, 2007) to estimate migration rates in each coastal pair, in allopatric comparisons between coastal populations inhabiting the same environment, and between four allopatric populations found in inland and alpine habitats. We expected to detect variable levels of gene flow between proximate (parapatric Dune and Headland pairs) but not geographically distant populations. We only detected asymmetric gene flow in two comparisons (from D03 to H02 and from H01 to H05), and failed to detect population migration in all other comparisons (Table 3.4, Table 3.5).

Overall, these results suggest that: 1) Dune and Headland ecotypes diverged recently; 2) geography influences overall patterns of genetic variability in *S. lautus*; 3) Dune and Headland populations diverged *in situ* and originated at least two times; 4) Dune and Headland populations are genetically similar to one another, and 5) Dune and Headland parapatric populations have experienced very little or no gene flow in the recent past (but note that gene flow was detected between two allopatric headland populations). Below we discuss these ideas and how they help us better understand local adaptation and its contribution to parapatric speciation.

DISCUSSION

Systems where the same traits conferring local adaptation evolve independently in closely related lineages might favour the study of how natural selection reduces gene flow between populations (Schluter & Nagel, 1995; Ostevik *et al.*, 2012). This is the case of coastal population pairs in *Senecio lautus*, where growth habits (e.g., prostrate versus erect) have evolved multiple times in contiguous populations inhabiting contrasting habitats along the coast. These traits have arisen independently at least two times, one in the eastern and one in the southern coasts of Australia (Fig. 3.1, Fig 2-3 in Roda *et al.* 2013). However, given the strong isolation by distance within each clade and the fact that populations cluster by geography and not by ecotype –regardless of whether neutral or outlier loci are used for phylogenetic reconstruction (Roda *et al.*, 2013a)– it is likely that the forms have evolved in parallel many times within each clade. Other studies in plants and animals have found trait evolution with phylogenetic independence, (Schluter & Nagel, 1995; Colosimo *et al.*, 2005; Baumbach & Hellwig, 2007; Foster *et al.*, 2007; Palkovacs *et al.*, 2008; Ostevik *et al.*, 2012; Strecker *et al.*, 2012), but only a few have found additional evidence for the parallel evolution of reproductive isolation between diverging populations (Schluter & Nagel, 1995;

Ostevik *et al.*, 2012). Below and in the following chapter, I provide insights as to whether this is happening in the *S. lautus* system and thus on how natural selection is shaping diversification in plants.

Reproductive isolation and gene flow in S. lautus

Dune and Headland population pairs from a single phylogenetic clade are genetically compatible in the glasshouse (F1 seed set is usually higher than parental seed set; Melo *et al.*, 2014 and Chapter IV) but exhibit strong reproductive isolation in the field (e.g., Melo *et al.*, 2014). Unsurprisingly, phylogenetic relationships show Dune and Headland populations have recently diverged, but persist differentiated in parapatry. Natural selection in the field is likely to be responsible for this differentiation, as transplant experiments have shown that both immigrants and F1 hybrids are selected against in the field (Melo *et al.*, 2014, Walter and Ortiz-Barrientos unpublished data). In terms of stages of speciation, this data is consistent with the origin of ecotypes in the system, but it does not help us to elaborate on whether the system is moving away from adaptive morphological differentiation to reproductive differentiation. Regardless, extant ecotype differentiation seems to have occurred in conditions that antagonised natural selection: not only are populations sister to each other, have biological traits consistent with high seed dispersal, and share pollinators, but their genomes have the signature of heterogeneous genomic divergence, one that becomes more accentuated as populations are further from one another (Roda *et al.*, 2013a). Therefore, molecular estimates of reproductive isolation, as measured by the rate of migration between populations, might help us to better understand whether natural selection has had some effect on the probability of gene exchange in parapatry. In particular, we expected that population pairs would be reproductively isolated as a function of their habitat differences (e.g., as a function of the difference in salt content between sand dunes and rocky headlands), and as a function of their distance between them (e.g., there would be more gene flow between a parapatric comparison than between an allopatric comparison).

We measured overall reproductive isolation between Dune and Headland pairs using coalescent approaches to measure neutral effective migration rate. We predicted that even if some neutral markers could flow easily between population pairs, they would do so as a function of the overall effect of natural selection on population divergence (Feder & Nosil,, 2010). This process, usually called Isolation by Adaptation (IBA; Nosil *et al.*, 2008), predicts that the strength of divergent natural selection modulates overall levels of gene flow between populations. Although we expected to perform this test by measuring the correlation between estimates of gene flow between populations and the environmental distance between their habitats, we were struck by the

unexpected results of lack of gene flow between most of the population comparisons in this study. In particular, we detected asymmetric gene flow between the parapatric pair at Cabarita Beach (*m*, from the Dune to the Headland), and between the allopatric Headland populations from Lennox Head and Coffs Harbour (separated for more than 200 km). This result contrasts with previous findings in the system where patterns of heterogeneous genomic divergence are consistent with episodes of gene flow in the *S. lautus* complex.

Heterogeneous genomic divergence results from the combined effects of drift and natural selection on allelic differentiation across the genome of two diverging populations. If populations are allopatric, theory predicts that basal genomic differentiation is high, and some areas of the genome evolving adaptively, or taken to fixation by drift, will be unusually differentiated between populations. On the other hand, when populations are parapatric, we expect that most of the genome will be similar between populations (i.e., basal genomic differentiation is low) and only those regions under strong selection will be differentiated –note that genetic drift can also affect divergence, but less so than in the allopatric case as most rare mutations unaffected by selection in parapatry will likely be lost before drift can take them to fixation in a population. The differential process between allopatry and parapatry leads to contrasting frequency distributions for Fst, for instance, where allopatric distributions will be flatter and more homogenous, and parapatric distributions will be more L-shaped and more heterogeneous.

Roda *et al.* (2013a) explored these patterns using genomic markers and found that as populations moved away from parapatry, Fst distributions where less L-shaped and became more homogenous. In other words, allopatric differentiation affected most of the genome, whereas parapatric differentiation affected small parts of the genome. Further, average Fst is lower in parapatry than in allopatry, and there is a strong positive correlation between Fst and geographic distance –usually the consequence of migration-drift equilibrium (Coyne & Orr, 2004)– suggesting that parapatric differentiation has occurred in the face of gene flow. So, how can we reconcile the results of Roda and colleagues (2013a) with our results of lack of gene flow, regardless of geography, reported here? We suggest that natural selection reduced gene flow between coastal populations in the recent past, and there has not been enough time to erode the signature of parapatric differentiation found by Roda *et al.* (2013a). In other words, we suggest that coastal ecotypes of *S. lautus* have differentiated first in the face of gene flow, and natural selection has recently created allopatric conditions for gene exchange between them, or an "allopatry-second" mode of speciation. Given that coalescent models of isolation with migration (IM and IMa2; Hey & Nielsen, 2004; Hey & Nielsen, 2007) could be inefficient at detecting gene flow at early stages of species divergence – possibly due to high variances in coalescence times of multiple loci (Becquet & Przeworski, 2009;

Strasburg & Rieseberg, 2011)– it is possible that levels of gene flow are just too low to be detected with the number of loci we used in this study; however, this does not affect our major contention that natural selection has played a major role in reducing gene flow in parapatric *Senecio*.

Understanding patterns of gene flow in parapatric Senecio

Our hypothesis is simple: Dune and Headland populations started diverging in the face of gene flow, which was subsequently precluded by strong ecology-based reproductive barriers. A major prediction of this dynamic is that ecology-based reproductive barriers will eventually facilitate the accumulation of patterns of genomic divergence similar to those found in populations diverging in allopatry. Below we discuss this scenario in detail, and test it by reanalysing the genome-wide genotypic data of Roda *et al*. (2013a). In particular, we focus on how patterns of shared polymorphism vary between parapatric and allopatric population pairs, while observing the genetic and geographic distance that separates them.

Figure 3.3 shows the theoretical predictions for geographic models of divergence based on the type of polymorphisms present between populations comparisons (Machado *et al.*, 2002). For a classic case of allopatric divergence (Fig. 3.3a), populations are expected to share some of the variation that was present in the ancestral populations (i.e., sites polymorphic in both populations, henceforth called PP sites), but with an excess of fixed mutations in one (FP of PF) or both populations (FF) – if Ne were particularly small–as a consequence of genetic drift over long periods of time and with some mutation input (Hartl & Clark, 1997). Without mutation, the population would eventually loose all of its variability, and the FF category would be the only one present in the system. For divergence in the presence of gene flow (Fig. 3.3b), we expect more variation to be shared between populations, creating an excess of shared neutral polymorphisms (PP) and a concomitant reduction of fixed polymorphic sites as they are transferred between populations via gene flow. In contrast, if populations started diverging in the presence of gene flow, but are currently isolated by strong RI (Fig. 3.3c), we expect to find an excess of shared polymorphism due to recent gene flow, but an increase in allelic fixation between populations due to genetic drift. Because most mutations will be unique, drift will mostly contribute to the FP and PF category, as predicted by the neutral theory (Kimura, 1985). Because of the presence of large fractions of both PP and FP and PF polymorphisms, we suggest that this reflects a mixed pattern of parapatric and allopatric genetic differentiation.

The *S. lautus* system includes proximate and distant populations, therefore we predict to find a negative correlation between the proportion of PP and geographic distance and a positive correlation between each of FP, PF, and FF with geographic distance. This is a decomposition of

isolation by distance, and perhaps the only unusual pattern would be that the intercept for FP and PF should be high and not close to zero, as it would be predicted for a purely case of neutral parapatric differentiation. In other words, average Fst in parapatry can be significantly different from zero, but it is expected to be low. We reanalysed the RADs data of Roda *et al.* (2013a) and estimated the proportion of each type of polymorphism across the whole genome of all possible proximate (parapatric Dune and Headland pairs and other populations separated by less than 100 km) and distant (all other pairs) comparisons between populations. We excluded from this analysis populations A05 and D32 for which RAD tags were not sequenced. Only one SNP per tag was used to avoid biased correlations amongst different SNPs within the same locus (note that we cannot correct for correlations due to linkage between loci as we do not have a genome sequence for the species). Because we are dealing with a multiple comparison testing framework, we used Mantel tests to calculate Pearson's correlation coefficients using the *vegan* package (Oksanen *et al.*, 2013) from R (R Core Team 2012). We built matrices for each polymorphism type, the log transformed geographic distance (LogGD) and divergence times (DT) estimated by *Beast. Because correlations were bounded between zero and one, and significance in permutations tests are given as the proportion of test-statistic values greater than the observed one, we recalculated significance values for negative correlations using minus (logGD).

Consistent with our predictions for a mixed model of speciation in the face of gene flow, or "allopatry second" resulting from strong natural selection in parapatry, we found that (1) the closer in geography populations were the more genetic variation they shared and the less exclusive variation they showed (Fig. 3.4), and (2) proximate populations displayed a substantial fraction of FF, FP and PF fixed sites (Fig. 3.5). For instance, all five papapatric pairs displayed high proportions of shared polymorphisms as denoted by the first five circles at the left of the PP panel (Fig. 3.5). In particular population pair H05 – D04 only separated by a few meters display the highest proportion of shared polymorphisms, as well as the population pair H02 – D03 for which we detected asymmetric gene flow for *m.* Population comparisons like A05 – D03, belonging to different geographic regions exhibited the smallest proportion of PP (at the bottom right side of the panel). Polymorphism patterns of fixed variation showed an asymmetry: for PF parapatric populations displayed low proportions of fixed variation, but in similar or higher proportion to distant allopatric populations from different geographic regions (e.g D01 – H21, D03 – H21 and H02 – H21). For example H01 – D01 pair had the highest proportion of FF and PF of the parapatric pairs, but also higher than for other allopatric comparisons. However, for FP three of the parapatric populations displayed the lowest proportion of FP, but the remaining two also show substantial proportion of this types of polymorphisms. This can bee seen by the grey circles at the left of each

of the panels, which correspond to the five parapatric pairs in the analysis; and the black circles at the right extreme of the x axis that correspond to distant population comparisons. Increased differentiation for parapatric pairs can be seen in the overlapping distribution for Fst values of neutral markers (Fig. S3.1).

Reconstructing the evolutionary history of Senecio lautus

Overall, our results suggest coastal populations of *S. lautus* have evolved by parallel speciation driven by natural selection, where each Dune and Headland pair experienced gene flow in the past but are currently isolated by strong reproductive barriers that create a complete absence of it. Below, we criticise this suggestion and ask what other alternative explanations can bare on our results.

Other forms of reproductive isolation

Although we have evidence for strong extrinsic reproductive isolation separating coastal populations of *S. lautus*, there could be other reproductive barriers acting in the system that we have not measured yet. For instance, conspecific pollen precedence (CPP), –a RI barrier found effective in the isolation of other plant species (Howard, 1999), could be a major process in gene flow between Dune and Headland populations. This barrier may evolve in response to gamete competition and perhaps decoupled from adaptation to environmental conditions. In other words, although ecotypic differentiation results from the effects of local adaptation on gene flow, the evolution of intrinsic reproductive isolation might have not evolved in response to natural selection. However, this scenario does not affect the mixed model of divergence, because (1) mutations for CPP may require low levels of gene flow to accumulate in parapatry (Thibert Plante & Hendry, 2010), or (2) CPP might have evolved directly in response to maladaptive hybridisation between locally adapted populations (Albert & Schluter, 2004; Hopkins, 2013). It will be interesting to measure this barrier in future studies, and in particular compare its strength in parapatric versus allopatry.

Allopatric differentiation between Dune and Headland populations

It is possible that Dune and Headland pairs have never been parapatric, and through out their history they have never exchanged genes. Furthermore, their ages suggest that they likely experienced periods of geographic separation during the latest glaciation cycles. If this were true, then high levels of the ancestral polymorphism we detected between proximate populations compared to the low levels found between distant comparisons may just reflect the very recent divergence of each population pair. In this case, we would have to argue that parallel evolution occurred multiple

times, in situ, and in the complete absence of gene flow (note that Dune and Headland populations are sister pairs and in the complete absence of gene flow, parallel evolution is the only explanation for such phylogenetic pattern). However, we believe this is an unlikely explanation because (1) there is a positive correlation between Fst and geographic distance, which is only predicted from migration-drift balance, and 2) seeds in the system are highly dispersible where they can migrate not only to the alternative environment but also to distant localities were other coastal pairs are found. Gene flow detected for Headland populations separated by ~170 Km supports this contention (see results section above).

We have under-sampled the distribution of migration rates across the genome

It is likely that estimates of gene flow based on ~22 markers (in each comparison) are insufficient to account for the overall pattern on gene exchange between two populations. More problematic, the markers we used could be linked to one another, thus further limiting the scope of our inferences. We were able to cross reference our DNA sequences with a recently produced linkage map in the Ortiz-Barrientos Laboratory (Roda, 2014) and discovered that seven DNA sequences were located in different chromosomes. Thus, our estimates of gene flow are derived from a data set with a weak correlation structure, thus suggesting that our inferences could be applied to the genome level. Further studies could increase sampling across the arms of most chromosomes, thus directly estimating the frequency distribution of genomic migration rates between population pairs. We expect to find large heterogeneity in this distribution, but overall, similar patterns of shared polymorphism as we detected in the data set of Roda *et al.* (2013a). Further, using more population pairs, and extra allopatric comparisons so we can avoid using multiple comparisons, we should be able to explore what regions of the genome show consistently low levels of migration rates between parapatric but not between allopatric population pairs. This will help us identify those genomic regions that might have responded to environmental conditions in similar fashion and thus contribute to our understanding of parallel adaptation and replicated evolution of traits.

The most likely explanation for our results is that gene flow accompanied the initial divergence of Dune and Headland populations which strong natural selection and forms of RI subsequently brought to a halt. The actual mechanisms of selection remain partially unknown because in the pairs where it has been measured, extrinsic reproductive isolation is not complete (Melo *et al.*, 2014). However, our data seems to suggest that local adaptation and the evolution of reproductive isolation has led to the complete cessation of gene flow in parapatry. There are two possible ways in which RI could lead to lack of gene flow anywhere in the genome between two populations: First, strong and direct selection favouring alternative alleles at each population (when $s>>m$) could lead to

complete RI between populations (DS; Feder *et al.*, 2012; Feder *et al.*, 2013). Second, gene flow accompanied the initial divergence of populations, but natural selection on multiple scattered loci rapidly created reductions of gene flow across entire genomes (GH; Feder *et al.*, 2012; Feder *et al.*, 2013). Our current data cannot distinguish these two possibilities, but preliminary results in our lab have found that QTL and association mapping for major morphological differences have a complex genetic basis (Roda 2014; Bernal and Ortiz-Barrientos unpublished results), possibly suggesting that adaptation has followed complex trajectories and required many mutations of small and moderate effect. Under these circumstances, there could have been enough time for selection and migration to coexist in the system, thus perhaps favouring a case of genome hitchhiking over direct selection in this system.

Our study contributes to the growing number cases that support the role of natural selection in driving the genome differentiation during speciation. For instance, stickleback fish found in multiple lake-stream systems show varying degrees of genomic differentiation, but in all cases driven by selection acting across large regions of the genome, and where variation in recombination rate seems to generate strong heterogeneous genomic divergence in each population pair (Roesti *et al.*, 2012). Similarly, several species of Sunflower seem to be genetically differentiated at similar genomic regions regardless of whether they are allopatric or parapatric populations with varying degrees of gene flow, possibly suggesting that selection has reduced effective migration rates across their genomes (Renaut *et al.*, 2013). Finally, the Z chromosome of hybridising *Ficedula* flycatchers appears to be more homogenously divergent than the rest of their genomes, indicating that perhaps this region has shifted from an early to a later phase of genomic divergence (Ellegren *et al.*, 2012; Feder *et al.*, 2012). Although we do not have data to test the effects of linkage on genetic differentiation, our results seem to suggest that if they were important, their role was only relevant during the early stages of genetic differentiation and adaptation, and that currently they are innocuous to the speciation process. This idea is not necessarily novel (c.f. Michel *et al.*, 2010), and it might indicate that the antagonism between natural selection and gene flow can be resolved, and that the effect of gene flow on divergence is particularly pronounce during the early stages of speciation with gene flow or the early stages of speciation by reinforcement following secondary contact.

Conclusion

Understanding the role that natural selection plays in driving speciation between parapatric populations remains a fundamental goal in evolutionary biology (Seehausen *et al.*, 2014). Phylogenetic patterns, estimates of gene flow between parapatric population pairs, and the

apportionment of genetic variability across geography suggested that coastal ecotypes of *S. lautus* have diverged recently in multiple sites, as previously found by Roda *et al.* (2013), and that gene flow was extensively reduced after their original split. We propose that strong extrinsic RI may have lead to the complete isolation of parapatric plant populations, thus facilitating further genome-wide neutral and adaptive differentiation and the progress towards speciation (Turelli *et al.*, 2001). Our results highlight how the interaction between natural selection and migration might override the initial effects of genetic linkage on patterns of divergence [although linkage itself may have evolved in response to migration-selection balance (Martin *et al.*, 2006)], and could imply that ecological barriers are fundamental to divergence and speciation in the *Senecio lautus* system.

Acknowledgements

This research was funded by the Australian Research Council grants to D.O-B. Maddie James conducted the laboratory work and bioinformatics. Huanle Liu provided advise in the bioinformatics. RADs sequences we used for one of our analyses were developed by Federico Roda. Diana Bernal, Melanie Wilkinson and Federico Roda provided assistance in the laboratory work. We thank the glasshouse services unit at the University of Queensland for assistance with pest control and glasshouse space maintenance throughout several stages of the experiments.

References

Albert AYK, Schluter D. 2004. Reproductive character displacement of male stickleback mate preference : reinforcement or direc selection? *Evolution* **58**: 1099-1107.

Ali S. 1966. Senecio lautus complex in Australia. III. The genetic system. *Australian Journal of Botany* **14**: 317-327.

Ali S. 1968. Senecio lautus complex in Australia. IV. The biology of the complex. *Phyton (Horn, Austria)* **13**: 53-62.

Ali SI. 1964. *Senecio lautus* complex in Australia. I. Taxonomic considerations and discussion of some of the related taxa from New Zealand. *Australian Journal of Botany* **12**: 282-291.

Ali SI. 1969. Senecio lautus complex in Australia. V. Taxonomic interpretations. *Australian Journal of Botany* **17**: 161-176.

Baumbach H, Hellwig F. 2007. Genetic differentiation of metallicolous and non-metallicolous *Armeria maritima* (Mill.) Willd. taxa (Plumbaginaceae) in Central Europe. *Plant Systematics and Evolution* **269**: 245-258.

Becquet C, Przeworski M. 2009. Learning about modes of speciation by computational approaches. *Evolution* **63**: 2547-2562.

Berglund ABN, Dahlgren S, Westerbergh A. 2004. Evidence for parallel evolution and site-specific selection of serpentine tolerance in *Cerastium alpinum* during the colonization of Scandinavia. *New Phytologist* **161**: 199-209.

Bolnick DI, Fitzpatrick BM. 2007. Sympatric speciation: models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst* **38**: 459-487.

Butlin RK, Galindo J, Grahame JW. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 2997-3007.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**: 1928-1933.

Coyne JA, Orr HA. 2004. *Speciation*: Sinauer Associates Sunderland, MA.

Dieckmann U, Doebeli M. 1999. On the origin of species by sympatric speciation. *Nature* **400**: 354-357.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**: 756-760.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**: 2611-2620.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. . *Evolutionary Bioinformatics Online* **1**.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**: 1567-1587.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**: 342-350.

Feder JL, Flaxman SM, Egan S, Comeault AA, Nosil P. 2013. Geographic Mode of Speciation and Genomic Divergence. *Annual Review of Ecology, Evolution, and Systematics* **44**: null.

Feder JL, Nosil P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**: 1729-1747.

Felsenstein J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**: 124-138.

Filchak KE, Roethele JB, Feder JL. 2000. Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature* **407**: 739-742.

Foster SA, McKinnon GE, Steane DA, Potts BM, Vaillancourt RE. 2007. Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. *New Phytologist* **175**: 370-380.

Fu Y-X, Li W-H. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**: 693-709.

Gavrilets S. 2003. Perspective: models of speciation: what have we learned in 40 years? *Evolution* **57**: 2197-2215.

Gavrilets S. 2004. Fitness landscapes and the origin of species. *Austral Ecology* **30**: 610-611.

Gavrilets S, Li H, Vose MD. 2000. Patterns of parapatric speciation. *Evolution* **54**: 1126-1134.

Harrison S, Rajakaruna N. 2011. *Serpentine: the evolution and ecology of a model system*: Univ of California Pr.

Hartl DL, Clark AG. 1997. *Principles of population genetics*: Sinauer associates Sunderland.

Hasegawa M, Kishino H, Yano T-a. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160-174.

Hatfield T, Schluter D. 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* **53**: 866-873.

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**: 570-580.

Hendry APHAP. 2009. Ecological speciation! Or the lack thereof? . *Canadian Journal of Fisheries and Aquatic Sciences* **66**: 1383-1398.

Hey J. 2010. Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**: 921-933.

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**: 747-760.

Hey J, Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 2785-2790.

Hird SM, Brumfield RT, Carstens BC. 2011. PRGmatic: an efficient pipeline for collating genome-enriched second-generation sequencing data using a 'provisional-reference genome'. *Molecular Ecology Resources* **11**: 743-748.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**: e1000862.

Hopkins R. 2013. Reinforcement in plants. *New Phytologist* **197**(4): 1095-1103.

Howard DJ. 1999. Conspecific sperm and pollen precedence and speciation. *Annual Review of Ecology and Systematics* **30**: 109-132.

Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome research* **9**: 868-877.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147-164.

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153-159.

Kay K, Whittall J, Hodges S. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology* **6**: 36.

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Research* **12**: 656-664.

Kimura M. 1985. *The neutral theory of molecular evolution*: Cambridge University Press.

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**: 2283-2285.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:**1451-1452.

Machado CA, Kliman RM, Markert JA, Hey J. 2002. Inferring the history of speciation from multilocus DNA sequence Data: The case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution* **19**: 472-488.

Martin G, Otto SP, Lenormand T. 2006. Selection for recombination in structured populations. *Genetics* **172**: 593-609.

Melo MC, Grealy A, Brittain B, Walter GM, Ortiz-Barrientos D. 2014. Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsell. *New Phytologist* **doi: 10.1111/nph.12779**.

Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL. 2010. Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences* **107**: 9724-9729.

Moonsamy PV, Bonella PL, Williams TC, Holcomb CL, Turenchalk GS, Blake LA, Hoglund BN, Rastrou M, Daigle DA, Simen BB. 2011. 201-P Use of the Fluidigm® access Array™ system provides simplified amplicon library preparation in next generation sequencing for high throughput HLA genotyping. *Human Immunology* **72**: S142.

Nei M. 1987. Molecular Evolutionary Genetics. **Columbia University Press, New York**.

Nosil P, Crespi BJ, Sandoval CP. 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* **417**: 440-443.

Nosil P, Egan S, Funk D. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes:" Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**: 316-336.

Nosil P, Vines T, Funk D. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**: 705-719.

Oksanen J, Blanchet F, Kindt R, Legendre P, Michin P, RB OH, Simpson G, Solymos P, Stevens M, Wagner H. 2013. Vegan: Community ecology package. . *R package version 2.0-6.* http://CRAN.R-project.org/package=vegan.

Ostevik KL, Moyers BT, Owens GL, Rieseberg LH. 2012. Parallel ecological speciation in plants? *International Journal of Ecology* **2012**.

Palkovacs EP, Dion KB, Post DM, Caccone A. 2008. Independent evolutionary origins of landlocked alewife populations and rapid parallel evolution of phenotypic traits. *Molecular Ecology* **17**: 582-597.

Perrier C, Bourret V, Kent MP, Bernatchez L. 2013. Parallel and non-parallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon. *Molecular Ecology* **22:** 5577-5593.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Powell THQ, Hood GR, Murphy MO, Heilveil JS, Berlocher SH, Nosil P, Feder JL. 2013. Genetic divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (Dipter: Tephritidae). *Evolution* **67**: 2561-2576.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.

Radford I, Cousens R, Michael P. 2004. Morphological and genetic variation in the *Senecio pinnatifolius* complex: are variants worthy of taxonomic recognition? *Australian Systematic Botany* **17**: 29-48.

Radford IJ, Cousens RD. 2000. Invasiveness and comparative life-history traits of exotic and indigenous *Senecio* species in Australia. *Oecologia* **125**: 531-542.

Renaut S, Grassa C, Yeaman S, Moyers B, Lai Z, Kane N, Bowers J, Burke J, Rieseberg L. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature communications* **4**: 1827.

Roda F. 2014. The genomic basis of parallel ecological speciation (Unpublished doctoral dissertation). *University of Queensland*.

Roda F, Ambrose L, Walter GM, Liu HL, Schaul A, Lowe A, Pelser PB, Prentis P, Rieseberg LH, Ortiz-Barrientos D. 2013. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Molecular Ecology* **22**: 2941-2952.

Roesti M, Hendry AP, Salzburger W, Berner D. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. *Molecular Ecology* **21**: 2852-2862.

Rundle HD, Nagel L, Boughman JW, Schluter D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**: 306.

Rundle HD, Nosil P. 2005. Ecological speciation. *Ecology Letters* **8**: 336-352.

Rundle HD, Whitlock MC. 2001. A genetic interpretation of ecologically dependent isolation. *Evolution* **55**: 198-201.

Schluter D. 2000. *Ecology of adaptive radiation*. Oxford University press.

Schluter D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* **16**: 372-380.

Schluter D, Conte G. 2009. Genetics and ecological speciation. *Proceedings of the National Academy of Sciences* **106**: 9955.

Schluter D, Nagel LM. 1995. Parallel Speciation by Natural Selection. *The American Naturalist* **146**: 292-301.

Schmieder R, Lim Y, Rohwer F, Edwards R. 2010. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* **11**: 341.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström Å. 2014. Genomics and the origin of species. *Nature Reviews Genetics* **15**: 176-192.

Strasburg JL, Rieseberg LH. 2011. Interpreting the estimated timing of migration events between hybridizing species. *Molecular Ecology* **20**: 2353-2366.

Strecker U, Hausdorf B, Wilkens H. 2012. Parallel speciation in *Astyanax* cave fish (Teleostei) in Northern Mexico. *Molecular Phylogenetics and Evolution* **62**: 62-70.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585-595.

Thibert Plante X, Hendry A. 2010. When can ecological speciation be detected with neutral loci? *Molecular Ecology* **19**: 2301-2314.

Thibert-Plante X, Hendry A. 2009. Five questions on ecological speciation addressed with individual-based simulations. *Journal of Evolutionary Biology* **22**: 109-123.

Thompson I. 2005. Taxonomic studies of Australian *Senecio* (Asteraceae): 5. The *S. pinnatifolius/S. lautus* complex. *Muelleria* **21**: 23-76.

Turelli M, Barton NH, Coyne JA. 2001. Theory and speciation. *Trends in Ecology & Evolution* **16**: 330-343.

Watterson G. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**: 256-276.

Zhai W, Nielsen R, Slatkin M. 2009. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Molecular Biology and Evolution* **26**: 273-283.

Tables

Table 3.1. Populations of *Senecio lautus* used in the study, and their localities.

Locality	Coordinates	Population ID
	S 28° 48' 22.10" E 153° 36' 9.94"	H1
Lennox Head (NSW)	S 28° 47' 10.7" E 153° 35'	D1
	S 28° 21' 45.07" E 150° 34' 46.82"	D2
Cabarita Beach (NSW)	S 28° 19' 54.66" E 153° 34' 17.04"	D3
	S 30° 18' 42.42" E 153° 8' 37.68"	H5
Coffs Harbour (NSW)	S 30° 18' 45.9" E 153° 08' 24.12"	D4
Portland, Cape Bridgewater (VIC)	S 38° 22' 49.6" E 141° 22' 07"	H12
Discovery Bay Costal Park (VIC)	S38° 19' 28.10" E141° 23' 42.80"	D32
	S 43° 11' 14.4" E 147° 50' 40.3"	H15
Porth Arthur (TAS)	S 43° 10' 33.0" E 147° 51' 16.0"	D14
	S 33° 09' 9.1" E 134° 15' 43.1"	H21
Point Labatt (SA)	S 33° 07' 30.9" E 134° 15' 57.0"	D23
Hebel (QLD)	S 28° 57' 46.80" E 147° 47' 51.72"	I01
Cameby (QLD)	S 26° 41' 45.70" E 150° 30' 31.22"	I02
Fall’s Creek (VIC)	S 36° 52' 21.5" E 147° 17' 19.5"	A03
Great Lake, Poatina Rd	S 41° 48' 33.1" E 146° 52' 13.9"	A05

Table 3.2. Summary of polymorphism statistics for each gene across populations in *Senecio lautus.* Mean and Standard Errors (S.E.) are reported for number of segregating sites (S), haplotype number (H), haplotype diversity (Hd), the estimate of 4Nu using the average number of nucleotide differences per site (π ; Nei, 1987), the Estimate of 4Nu per base pair using the number of polymorphic sites (Theta; Watterson, 1975) and for estimates of Tajima's D.

	S		**H**		**Hd**		**π**		**Theta**		**D**	
Loci	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
1004	4.000	0.658	3.625	0.352	0.520	0.056	0.004	0.002	0.004	0.001	-0.862	0.193
1011	3.063	0.854	3.438	0.707	0.314	0.072	0.002	0.000	0.003	0.001	-1.028	0.221
1014	3.813	0.607	4.188	0.502	0.367	0.054	0.002	0.000	0.003	0.000	-1.257	0.145
1018	5.875	1.169	5.000	0.791	0.538	0.078	0.003	0.001	0.005	0.001	-1.194	0.221
1021	10.563	1.628	5.875	0.612	0.573	0.061	0.005	0.001	0.008	0.001	-1.493	0.197
1024	3.063	0.581	3.375	0.446	0.308	0.046	0.001	0.000	0.002	0.000	-1.152	0.191
1080	5.438	1.497	4.125	0.598	0.519	0.060	0.002	0.000	0.004	0.001	-0.838	0.216
1084	3.875	0.785	3.875	0.598	0.457	0.065	0.002	0.000	0.003	0.001	-1.048	0.164
1085	5.125	1.390	3.125	0.272	0.379	0.051	0.002	0.001	0.005	0.001	-1.093	0.261
1090	5.688	1.094	3.688	0.445	0.381	0.064	0.002	0.000	0.004	0.001	-1.434	0.228
1091	6.438	0.780	5.000	0.524	0.652	0.052	0.005	0.001	0.006	0.001	-0.441	0.245
1096	3.688	0.845	3.125	0.523	0.498	0.083	0.003	0.001	0.004	0.001	-0.407	0.245
1098	6.125	0.811	5.750	0.609	0.524	0.060	0.002	0.000	0.007	0.003	-1.710	0.098
1116	2.125	0.676	2.125	0.315	0.129	0.024	0.001	0.000	0.002	0.001	-1.535	0.104

Continuation Table 3.2. Summary of polymorphism statistics for each gene across populations in *Senecio lautus.* Mean and Standard Errors (S.E.) are reported for number of segregating sites (S), haplotype number (H), haplotype diversity (Hd), the estimate of 4Nu using the average number of nucleotide differences per site (π ; Nei, 1987), the Estimate of 4Nu per base pair using the number of polymorphic sites (Theta; Watterson, 1975) and for estimates of Tajima's D

	S		**H**		**Hd**		**π**		**Theta**		**D**	
Loci	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
1125	8.750	1.672	5.063	0.512	0.641	0.053	0.004	0.001	0.007	0.001	-1.235	0.170
1126	3.875	0.912	3.125	0.473	0.333	0.057	0.002	0.000	0.003	0.001	-1.368	0.133
1170	2.438	0.626	2.563	0.387	0.358	0.068	0.002	0.000	0.002	0.001	-0.756	0.211
1174	6.063	0.849	5.125	0.455	0.552	0.047	0.003	0.000	0.005	0.001	-1.217	0.125
1176	2.563	0.563	2.938	0.295	0.398	0.043	0.001	0.000	0.002	0.000	-0.625	0.204
1192	3.875	0.688	3.688	0.463	0.457	0.063	0.002	0.000	0.003	0.001	-0.591	0.259
1194	4.000	1.176	2.813	0.332	0.281	0.042	0.001	0.000	0.003	0.001	-0.992	0.333
1199	11.563	2.125	6.000	0.645	0.673	0.063	0.009	0.002	0.011	0.002	-1.279	0.169
1204	1.813	0.306	2.688	0.416	0.388	0.059	0.001	0.000	0.002	0.001	-0.671	0.243
1205	3.625	0.638	3.500	0.408	0.414	0.049	0.002	0.000	0.003	0.001	-1.354	0.111
1211	3.188	0.614	3.563	0.508	0.274	0.046	0.002	0.000	0.003	0.001	-1.389	0.103
1212	2.063	0.370	2.750	0.413	0.218	0.044	0.001	0.000	0.001	0.000	-1.158	0.124

Table 3.3. Tajima's D values for 26 nuclear genes for 16 populations of *Senecio lautus* in Australia. Genes in red are selected genes, genes that exhibited recombination events are in blue, and np correspond to non-polymorphic.

Locus	A03	A05	I01	I02	D1	D3	D4	D14	D23	D32	H1	H2	H5	H12	H15	H21
1004	-1.595	-1.034	-1.233	-1.359	-0.521	-0.741	np	-0.030	-1.667	-1.959	0.458	-0.438	-1.959	-0.532	-0.117	-0.195
1011	-0.592	np	-1.762	-1.367	-1.508	np	-1.162	np	-1.088	-0.883	-0.354	-1.164	-1.515	-1.688	-1.508	1.232
1014	-1.096	-1.515	-1.780	-1.638	-0.494	-0.641	-1.667	-1.733	-1.573	-1.667	0.166	-0.903	-0.892	-1.997	-1.162	-1.524
1018	-0.057	np	0.888	-1.838	np	-0.887	-1.038	-1.471	-1.780	-2.144	-1.854	-1.401	-1.987	-0.909	-0.809	-1.434
1021	-2.097	-1.703	-1.875	-1.207	-2.208	-1.638	-2.076	-2.105	**np**	-1.209	0.107	-1.772	-2.300	-1.336	-1.127	0.151
1024	-1.729	-1.471	-1.667	-1.515	-0.975	0.895	-0.592	np	-1.086	-2.241	-0.382	-1.498	-1.162	-1.682	-0.835	-1.337
1080	-0.537	-1.865	-1.908	-0.632	-1.055	-0.133	-2.281	0.285	-0.249	-1.479	np	-0.369	np	-0.768	0.273	-1.019
1084	-1.027	-0.981	-1.349	-1.527	np	-0.195	-2.003	-0.195	-1.401	-1.190	-1.959	-0.809	np	-1.155	0.019	-0.905
1085	0.334	-2.145	-2.085	-1.508	-1.907	-1.072	-1.713	-1.278	-1.929	-1.878	0.671	1.166	-0.681	-1.723	-1.498	-0.248
1090	-1.401	-1.629	-1.421	1.167	-1.443	-2.098	-2.316	-1.671	-1.878	-1.090	-1.515	-2.003	np	-0.212	-1.831	-2.175
1091	-0.841	-0.857	-0.574	0.586	-0.561	0.024	-0.920	0.828	-0.729	-1.451	-1.448	-0.196	1.740	-1.889	0.578	-1.338
1096	1.167	-1.155	-1.233	-0.745	-0.729	-0.976	-1.259	np	np	-0.959	-0.013	0.889	-0.482	1.032	np	-0.829
1098	-1.779	-1.762	-2.213	-1.826	-2.107	-1.878	-0.993	-1.363	-1.891	-1.387	-1.515	-2.003	np	-1.733	-1.042	-2.154
1116	-1.956	-1.165	np	-1.723	-1.513	-1.508	np	-1.162	-1.162	-1.162	-2.333	np	-1.868	-1.515	-1.723	-1.159
1125	-1.136	-1.141	-0.329	-0.783	-0.812	-1.929	-1.839	-1.630	-1.452	-1.943	-1.521	-1.779	0.334	-2.089	-0.459	-1.245
1126	-1.034	-1.741	0.015	-1.385	-1.310	-1.141	-1.307	-1.162	-1.187	-2.225	-1.697	np	-1.798	-1.562	-1.111	-1.881
1170	0.429	-1.112	-1.877	-1.839	-0.661	np	0.334	-1.155	-1.141	-0.562	0.324	-0.959	-0.448	np	np	-1.165

Continuation Table 3.3. Tajima's D values for 26 nuclear genes for 16 populations of *Senecio lautus* in Australia. Genes in red are selected genes, genes that exhibited recombination events are in blue, and np correspond to non-polymorphic.

Locus	A03	A05	I01	I02	D1	D3	D4	D14	D23	D32	H1	H2	H5	H12	H15	H21
1174	-1.939	-1.141	-1.632	-0.414	-1.183	-0.414	-1.131	-0.901	-0.706	-1.823	-1.780	-1.498	-1.851	-0.788	-1.273	-0.992
1176	-1.451	-1.141	-1.112	-1.792	-1.179	-1.385	-0.729	-0.178	-0.035	-0.248	-0.195	-1.451	-0.195	-0.195	-0.195	1.486
1192	0.414	np	-1.092	-1.498	0.592	-0.592	-1.087	-1.650	-0.654	-1.315	0.677	-1.165	1.659	-0.438	-0.848	-1.865
1194	-1.729	-1.509	-1.729	-2.178	1.026	1.334	-0.641	1.596	-1.495	-1.513	-0.681	-1.162	-2.227	np	-1.729	-2.245
1199	-1.357	-0.728	-0.581	-1.292	np	-1.715	-1.610	-1.386	-1.750	0.319	-1.907	-1.193	-0.476	-2.156	-1.652	-1.697
1204	-1.141	-1.527	-0.448	-1.723	-0.850	-1.481	1.381	-0.195	-0.850	-0.850	0.019	1.212	-0.859	-1.451	-1.310	np
1205	-1.747	-1.717	-1.959	-0.565	-1.401	np	-0.959	-1.315	-0.448	-1.310	-1.385	-1.401	-1.713	-1.189	-1.747	-1.451
1211	-1.161	-1.667	-1.515	-1.997	-1.893	-0.603	-1.394	-1.041	-1.515	-1.884	-1.162	-1.250	-1.212	-1.159	np	np
1212	np	-1.162	-0.641	np	-1.202	-0.920	-1.515	-1.515	-1.682	-1.515	-0.283	-1.733	-1.023	-0.607	np	-1.256

Table 3.4. Maximum-Likelihood Estimates (MLE) and Highest Posterior Density Intervals (HPD) for population parameters for parapatric populations. Terms **q0**, **q1** and **q2** correspond to population sizes of population 1, population 2 and ancestral population, **m1>0** migration rate from population 1 to population 2 in the comparison, **m0>1** migration rate from population 2 to population 1, to divergence time, **N0, N1** and **N2** correspond to effective population sizes of population 1, population 2 and ancestral population **2N1m1>0** population migration rate from population 1 to population 2, **2N0m0>1** population migration rate from population 2 to population 1, and **T** to divergence time in years.

Comparison	**q0**	**q1**	**q2**	**m1>0**	**m0>1**	**to**	**N0**	**N1**	**N2**	**2N1m1>0**	**2N0m0>1**	**T**
A03-I02												
HtPt	1.522	2.37	0.294	0.2355	0.0405	0.1815	78530	122283	15169	0.4348	0.02699	37459
HPD95Lo	0.666	1.162	0.126	0	0	0.0765	34363	59955	6501	0	0	15788
HPD95Hi	3.214	3.894	0.61	2.398	2.296	0.2835	165831	200916	31474	2.72	1.82	58510
A05-I01												
HtPt	0.554	1.054	0.018	0.0645	0.7425	0.0915	33898	64492	1101	0.05097	0.2489	22395
HPD95Lo	0.178	0.374	0	0	0	0.0345	10891	22884	0	0	0	8444
HPD95Hi	1.558	2.782	0.09	2.63	2.74	0.1755	95330	170224	5507	1.964	1.142	42954
H01-D01												
HtPt	0.895	1.135	0.695	0.0015	0.0015	0.1185	34380	43600	26698	0.006836	0.002186	18208
HPD95Lo	0.495	0.545	0.435	0	0	0.0645	19015	20936	16710	0	0	9911
HPD95Hi	1.535	2.805	1.045	2.14	2.675	0.2055	58965	107751	40142	1.538	1.235	31576

Continuation Table 3.4. Maximum-Likelihood Estimates (MLE) and Highest Posterior Density Intervals (HPD) for population parameters for parapatric population pairs. Terms **q0**, **q1** and **q2** correspond to population sizes of population 1, population 2 and ancestral population, **m1>0** migration rate from population 1 to population 2 in the comparison, **m0>1** migration rate from population 2 to population 1, **to** divergence time, **N0**, **N1** and **N2** correspond to effective population sizes of population 1, population 2 and ancestral population **2N1m1>0** population migration rate from population 1 to population 2, **2N0m0>1** population migration rate from population 2 to population 1, and **T** to divergence time in years.

Comparison	**q0**	**q1**	**q2**	**m1>0**	**m0>1**	**to**	**N0**	**N1**	**N2**	**2N1m1>0**	**2N0m0>1**	**T**
H02-D03												
HtPt	1.035	0.165	0.315	3.342	**23.01**	0.075	51084	8144	15547	0.5002	9.708	14807
HPD95Lo	0.285	0.075	0.045	0	**12.13**	0	14067	3702	2221	0	0	0
HPD95Hi	25.52	0.525	27.77	23.71	**34.98**	29.98	1259338	25912	1370391	1.51	333.5	59198
H05-D04												
HtPt	0.315	0.225	0.555	0.945	0.015	0.525	23906	17075	42120	0.2247	0.2249	159371
HPD95Lo	0.075	0.105	0.165	0	0	0.045	5692	7969	12522	0	0	13660
HPD95Hi	8.055	0.885	29.98	26.8	29.98	1.215	611302	67164	2275592	6.965	44.75	368830
H12-D32												
HtPt	0.5125	4.487	0.5375	0.405	0.015	0.075	17491	153149	18344	1.312	1.687	10238
HPD95Lo	0.1125	1.663	0.2625	0	0	0.035	3839	56738	8959	0	0	4778
HPD95Hi	10.86	23.09	0.9375	18.38	26.11	0.195	370714	787927	31995	125.7	36.17	26620

Continuation Table 3.4. Maximum-Likelihood Estimates (MLE) and Highest Posterior Density Intervals (HPD) for population parameters for parapatric population pairs. Terms **q0**, **q1** and **q2** correspond to population sizes of population 1, population 2 and ancestral population, **m1>0** migration rate from population 1 to population 2 in the comparison, **m0>1** migration rate from population 2 to population 1, **to** divergence time, **N0**, **N1** and **N2** correspond to effective population sizes of population 1, population 2 and ancestral population **2N1m1>0** population migration rate from population 1 to population 2, **2N0m0>1** population migration rate from population 2 to population 1, and **T** to divergence time in years.

Comparison	**q0**	**q1**	**q2**	**m1>0**	**m0>1**	**to**	**N0**	**N1**	**N2**	**2N1m1>0**	**2N0m0>1**	**T**
H15-D14												
HtPt	0.2585	0.1631	0.123	0.025	0.525	0.0238	24012	15153	11423	0.04408	0.06272	8844
HPD95Lo	0.05271	0.04769	0.01255	0	0	0.003105	4896	4429	1166	0	0	1154
HPD95Hi	1.614	0.6702	2.573	30.68	43.12	2.069	149902	62246	238958	3.13	11.35	76866
H21-D23												
HtPt	2.145	1.865	0.285	0.03	0.09	0.08441	154089	133975	20473	0.9486	0.04997	24255
HPD95Lo	0.735	0.585	0.105	0	0	0.03353	52800	42024	7543	0	0	9635
HPD95Hi	6.845	6.575	0.525	14.07	10.25	0.1469	491721	472325	37714	15.03	14.64	42197

Table 3.5. Maximum-Likelihood Estimates (MLE) and Highest Posterior Density Intervals (HPD) for population parameters for allopatric population pairs. Terms **q0**, **q1** and **q2** correspond to population sizes of population 1, population 2 and ancestral population, **m1>0** migration rate from population 1 to population 2 in the comparison, **m0>1** migration rate from population 2 to population 1, **to** divergence time, **N0**, **N1** and **N2** correspond to effective population sizes of population 1, population 2 and ancestral population **2N1m1>0** population migration rate from population 1 to population 2, **2N0m0>1** population migration rate from population 2 to population 1, and **T** to divergence time in years.

Comparison	**q0**	**q1**	**q2**	**m1>0**	**m0>1**	**to**	**N0**	**N1**	**N2**	**2N1m1>0**	**2N0m0>1**	**T**
H01-H05												
HtPt	0.225	0.225	0.465	**4.695**	0.135	0.585	8480	8480	17526	**0.8927**	0.03094	88197
HPD95Lo	0.105	0.105	0.165	**0.705**	0.0	0.045	3958	3958	6219	**0.05434**	0.0	6784
HPD95Hi	0.675	0.705	28.04	**15.59**	5.445	29.98	25441	26572	1056665	**1.736**	0.5655	4520650
H15-H21												
HtPt	0.645	1.635	0.075	0.015	0.015	0.135	30325	76871	3526	0.01001	0.002138	25389
HPD95Lo	0.255	0.645	0.0	0.0	0.0	0.045	11989	30325	0.0	0.0	0.0	8463
HPD95Hi	1.365	3.465	0.4050	1.635	3.135	0.435	64177	162910	19041	1.372	0.9127	81808
H12-H15												
HtPt	0.195	0.255	29.98	0.045	0.015	0.495	12495	16339	1921274	0.01159	0.08291	126868
HPD95Lo	0.105	0.105	16.84	0.0	0.0	0.225	6728	6728	1079335	0.0	0.0	57667
HPD95Hi	0.645	0.675	29.98	29.14	28.57	1.275	41328	43250	1921274	5.481	4.726	326780

Table 3.5. Maximum-Likelihood Estimates (MLE) and Highest Posterior Density Intervals (HPD) for population parameters for allopatric population pairs. Terms **q0**, **q1** and **q2** correspond to population sizes of population 1, population 2 and ancestral population, **m1>0** migration rate from population 1 to population 2 in the comparison, **m0>1** migration rate from population 2 to population 1, **to** divergence time, **N0**, **N1** and **N2** correspond to effective population sizes of population 1, population 2 and ancestral population **2N1m1>0** population migration rate from population 1 to population 2, **2N0m0>1** population migration rate from population 2 to population 1, and **T** to divergence time in years.

Comparison	q0	q1	q2	m1>0	m0>1	to	N0	N1	N2	2N1m1>0	2N0m0>1	T
H01-H02												
HtPt	0.465	0.465	0.585	0.015	0.075	0.105	20481	20481	25767	0.2249	0.1097	18499
HPD95Lo	0.195	0.195	0.255	0.0	0.0	0.045	8589	8589	11232	0.0	0.0	7928
HPD95Hi	0.975	1.125	29.98	14.05	14.03	29.95	42945	49552	1320721	5.622	3.620	5277597
D14-D32												
HtPt	0.165	2.535	0.615	0.015	0.015	0.225	7132	109579	26584	0.1203	0.001238	38904
HPD95lo	0.075	1.185	0.1950	0.0	0.0	0.105	3242	51223	8429	0.0	0.0	18155
HPD95Hi	1.155	23.05	29.98	9.495	7.695	2.265	49927	996586	1296145	88.87	0.6299	391632
D14-D23												
HtPt	0.375	1.095	0.255	0.015	0.285	0.135	28469	83130	19359	0.04646	0.07076	40996
HPD95Lo	0.135	0.375	0.0	0.0	0.0	0.045	10249	28469	0.0	0.0	0.0	13665
HPD95Hi	0.945	2.745	22.73	14.60	4.965	29.89	71743	208395	1725238	7.109	0.8702	9078283

Continuation Table 3.5. Maximum-Likelihood Estimates (MLE) and Highest Posterior Density Intervals (HPD) for population parameters for allopatric population pairs. Terms **q0**, **q1** and **q2** correspond to population sizes of population 1, population 2 and ancestral population, **m1>0** migration rate from population 1 to population 2 in the comparison, **m0>1** migration rate from population 2 to population 1, **to** divergence time, **N0**, **N1** and **N2** correspond to effective population sizes of population 1, population 2 and ancestral population **2N1m1>0** population migration rate from population 1 to population 2, **2N0m0>1** population migration rate from population 2 to population 1, and **T** to divergence time in years.

Comparison	**q0**	**q1**	**q2**	**m1>0**	**m0>1**	**to**	**N0**	**N1**	**N2**	**2N1m1>0**	**2N0m0>1**	**T**
D01-D04												
HtPt	0.525	0.465	0.555	1.245	0.105	0.075	24530	21727	25932	0.6747	0.2249	14017
HPD95Lo	0.135	0.075	0.1950	0.0	0.0	0.0	6308	3504	9111	0.0	0.0	0.0
HPD95Hi	21.89	4.695	26.59	25.75	28.82	29.89	1022549	219368	1242618	28.56	198.1	5587224
D01-D03												
HtPt	1.425	0.375	0.135	0.555	0.795	0.075	75993	19998	7199	2.921	0.2249	15999
HPD95Lo	0.165	0.105	0.0	0.0	0.0	0.0	8799	5599	0.0	0.0	0.0	0.0
HPD95Hi	23.66	1.725	3.555	27.68	29.98	29.98	1261483	91991	189582	11.01	230.5	6396206

Figures

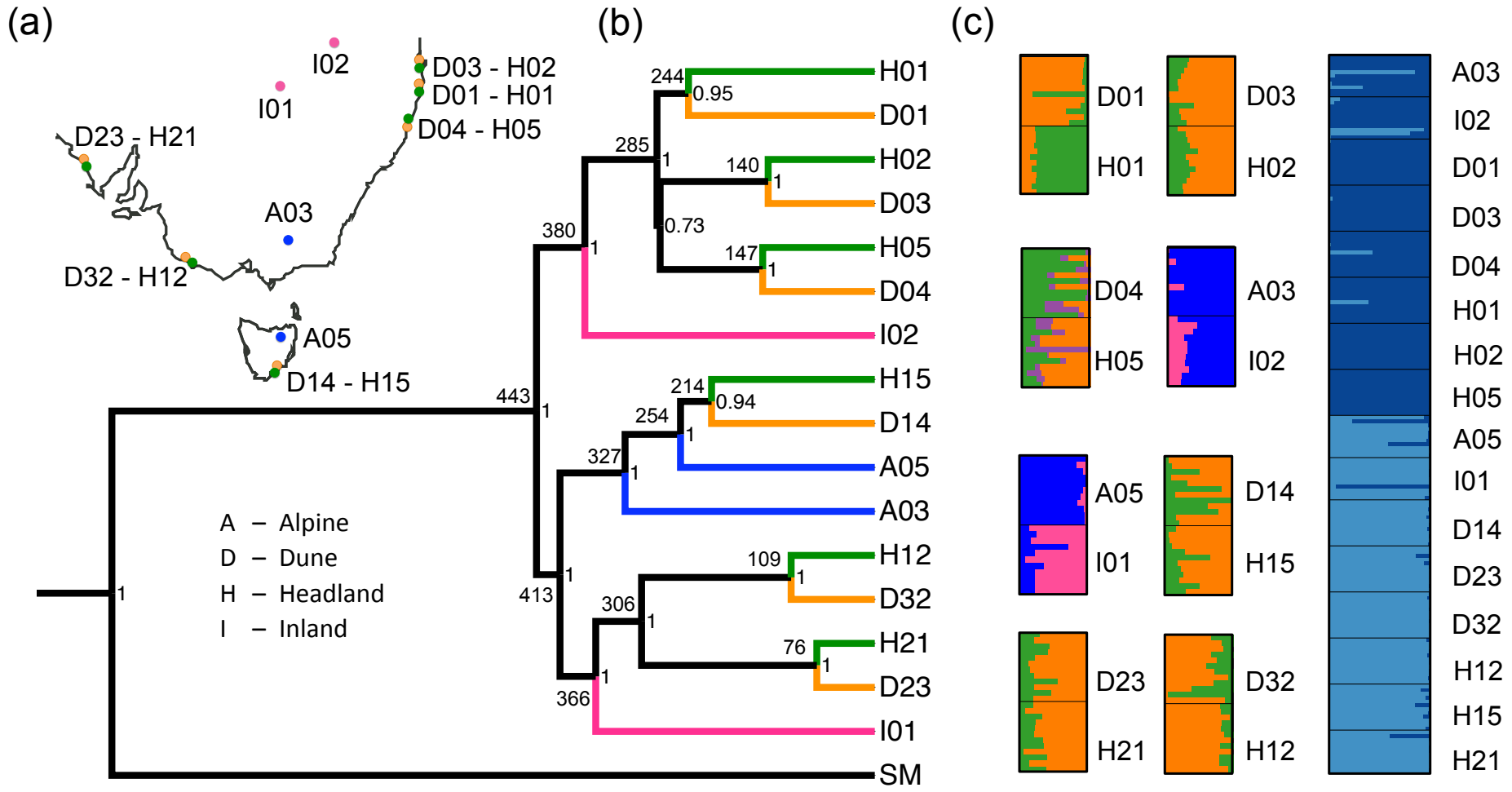

Fig. 3.1. a) Geographic distribution of *Senecio lautus* populations in the study. b) Phylogeny based on 13 neutral markers using Bayesian inference of the populations in the study. c) STRUCTURE analysis genetic clusters for all populations in the study and by pairs.

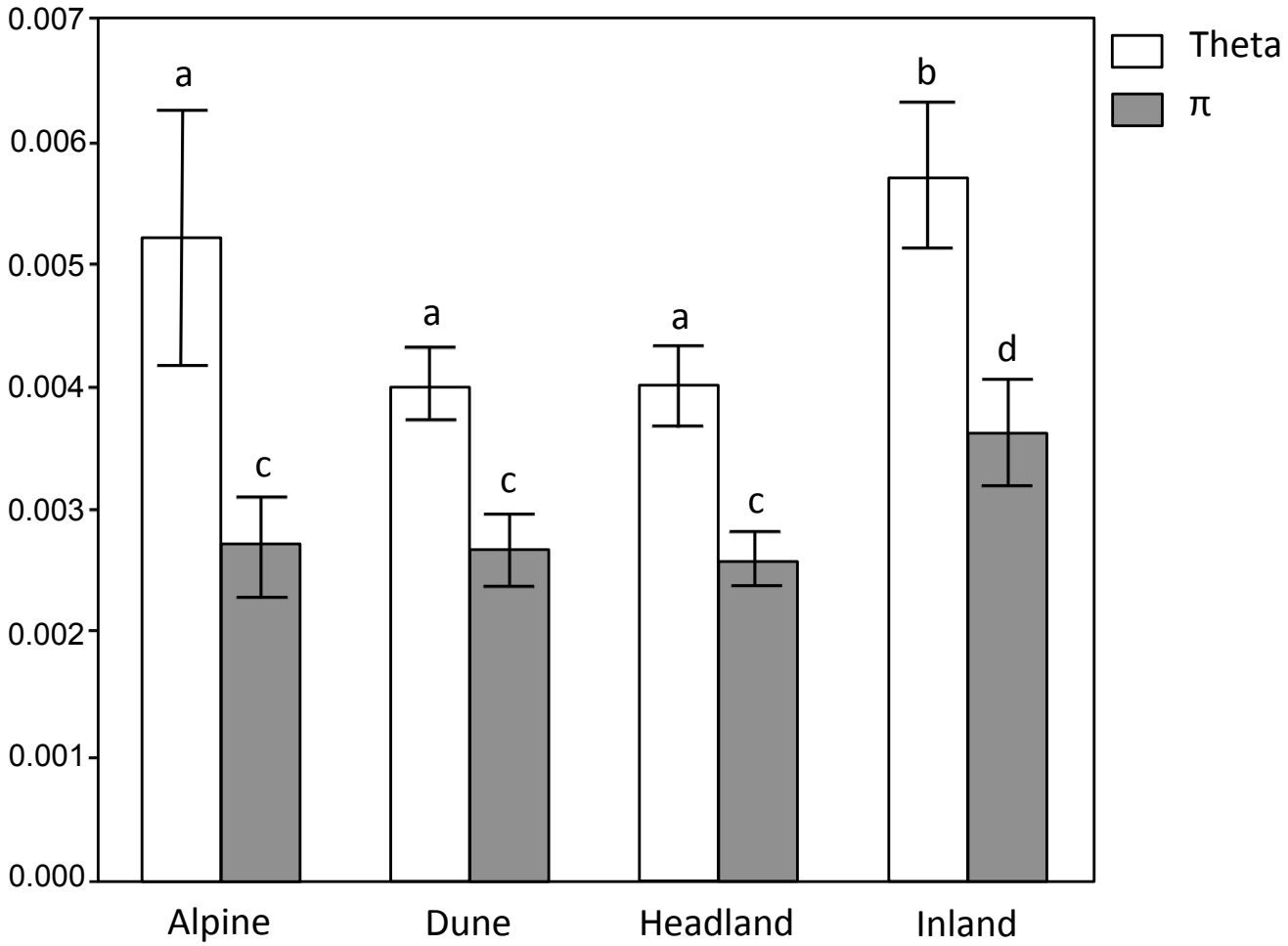

Fig. 3.2. Average π and Theta in the polymorphism analysis for ecotypes in *Senecio lautus*.

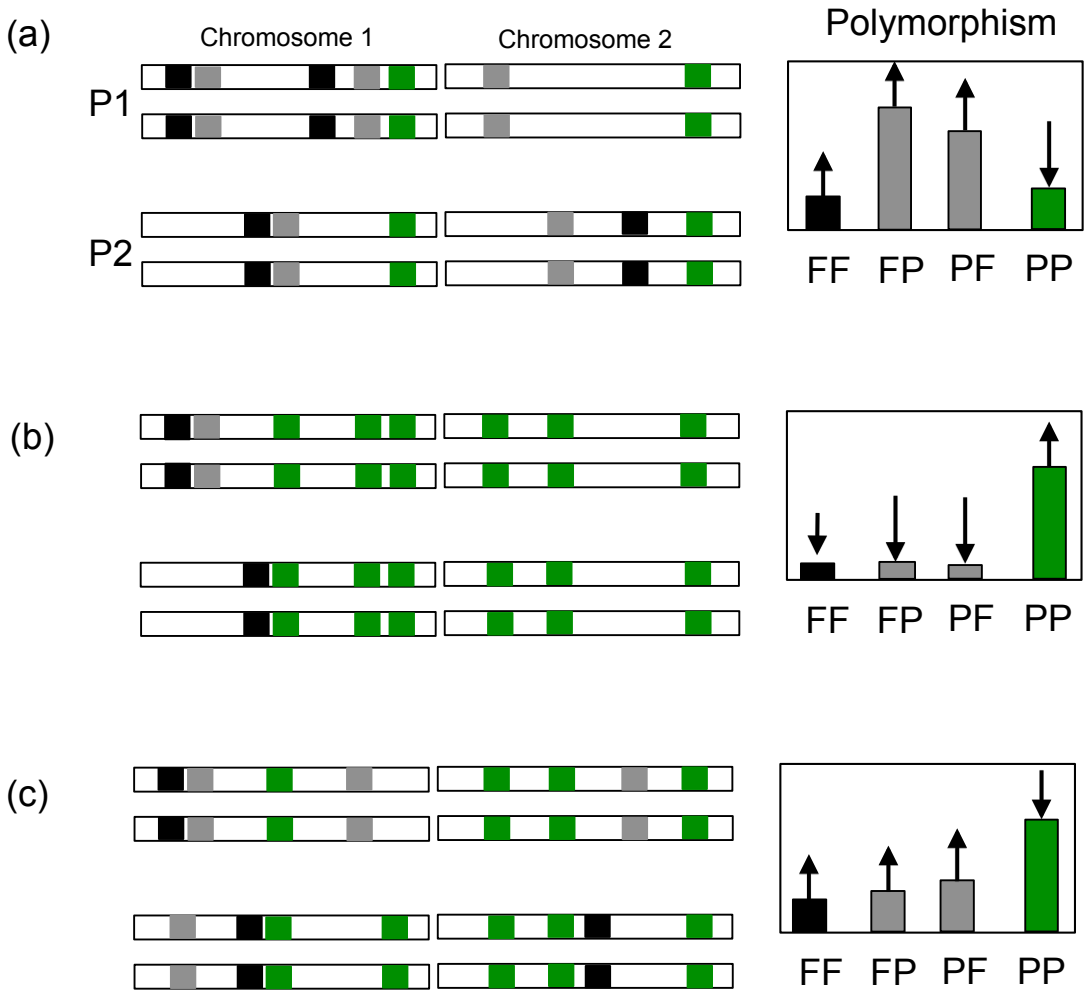

Fig. 3.3. Predictions for polymorphism patterns when population of are diverging in a) allopatry b) in the face of gene flow c) with gene flow but widespread effect of selection. FF (fixed-fixed), FP, PF (fixed in one population and polymorphic in the other one), and PP (polymphic-polymorphic).

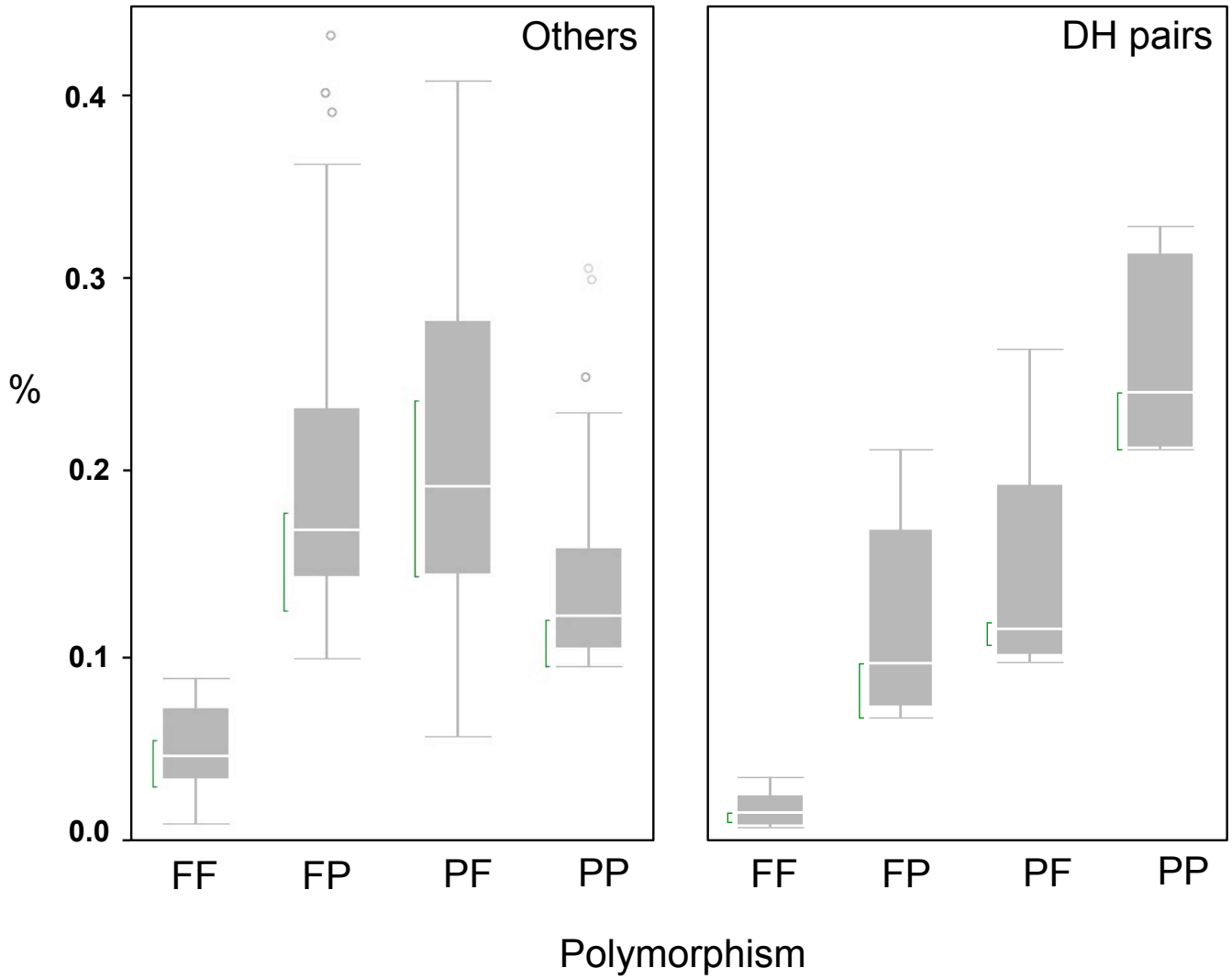

Fig. 3.4. Proportion of FF (fixed-fixed), FP, PF (fixed in one population and polymorphic in the other one), and PP (polymphic-polymorphic), polymorphisms in *Senecio lautus,* between Dune and Headland parapatric comparisons (left panel) and other allopatric comparisons (right panel).

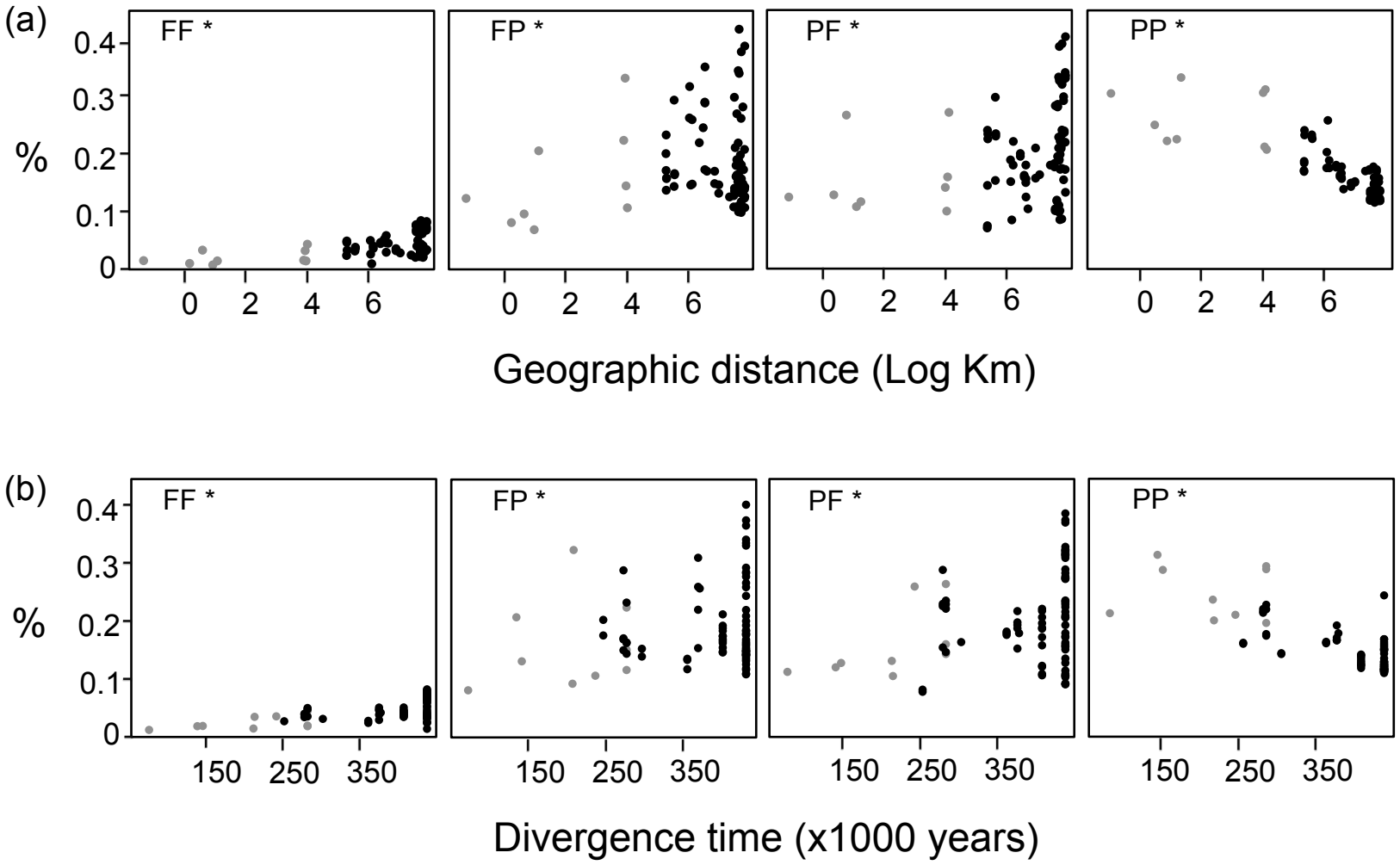

Fig. 3.5. Polymorphism patterns for populations pairwise comparisons in *Senecio lautus* across a) log geographic distance and b) divergence times. We found a positive correlation between the proportion of pairwise fixed and fixed-polymorphic sites with geographic distance between populations (FF: $r = 0.5374$, $p = 0.001$; FP: $r = 0.1424$, $p = 0.085$ and PF: $r = 0.3023$, $p = 0.005$). Contrastingly the proportion of shared polymorphic sites between pairs decreases with distance between them (PP: $r = -0.8266$, $p = 0.001$). The same pattern was found for each

polymorphism while correcting for divergence time between population pairs (FF: $r = 0.598$, $p = 0.001$; FP: $r = 0.1885$, $p = 0.015$ and PF: $r = 0.3111$, $p = 0.001$, and PP: $r = -0.7858$, $p = 0.001$). Although rare alleles may contribute to an increase of polymorphic-fixed sites even when there is partial gene flow between populations, we did not find any changes to our results when we repeated the analyses excluding all alleles with frequency below five or ten percent in any population.

CHAPTER IV

PARALLEL ECOLOGICAL SPECIATION IN *SENECIO LAUTUS*

ABSTRACT

Ecological parallel speciation or the repeated evolution of populations with the same mechanisms of reproductive isolation could provide the strongest evidence for the role of natural selection in speciation. Parallel ecological speciation systems must meet several conditions (independence, selection, compatibility, and isolation; Ostevik *et al*., 2012) that altogether rule out other possible models of speciation. Coastal populations in *Senecio lautus* have strong evidence for the *independence* of Dune and Headland populations, and several transplant experiments suggest that predation and edaphic differences in the soil (e.g. salt) are sources of strong natural selection, *selection*. Here I test for the other two remaining conditions of the model of parallel ecological speciation: reproductive *isolation* between populations adapted to different habitats, and reproductive *compatibility* between populations adapted to the same habitat. I based the test on three reproductive barriers (one extrinsic and two intrinsic), and included geographic region to test for the role of allopatry in the evolution of reproductive isolation. I discovered that consistent with the prediction of parallel ecological speciation, Dune and Headland populations are more reproductively isolated than Dune – Dune and Headland – Headland comparisons, with a stronger trend when comparisons included populations from different regions. Our results provide strong evidence for the parallel ecological speciation in *S. lautus,* and suggest that intrinsic reproductive isolation barriers are more likely to arise long after lineages have ecologically diverged.

Key words: Ecological parallel speciation, intrinsic postzygotic isolation, extrinsic postzygotic isolation, gene flow, parapatry, immigrant inviability, F1 seed set, hybrid viability.

INTRODUCTION

Ecological speciation with gene flow or the evolution of reproductive isolation (RI) between populations adapting to contrasting environments is one the most intricate processes in evolution (Coyne & Orr, 2004; Nosil, 2008). Although traditionally studies of speciation have focused on dissecting the reproductive isolating barriers that arise along the process (Coyne & Orr, 1998; Coyne & Orr, 2004), it remains unclear whether these barriers contributed directly to the origin of the species under study. This may be the case in studies of intrinsic postzygotic RI, where the evolution of genetic incompatibilities can occur on already formed species so the genetic basis of hybrid sterility or inviability may reflect post-speciation dynamics which are not guaranteed to be the same as those occurring during earlier stages of speciation (Agrawal *et al.*, 2011; Nosil, 2012). With out better luck, studies of ecological speciation using young systems that have not evolved complete RI could face similar uncertainty: recently diverged populations can bias the study of speciation towards extrinsic reproductive barriers that promote the formation of ecotypes, but that might not drive the completion of speciation (Coyne & Orr, 2004; Via, 2009). One scenario that dissipates doubts about the role of RI barriers to the process of speciation is in cases of parallel ecological speciation where there is a direct link between extrinsic and intrinsic RI and the tempo of divergence falls within the speciation continuum (Schluter & Nagel, 1995; Nosil, 2012; Ostevik *et al.*, 2012).

Ecological parallel speciation is one of the strongest predictions for ecological speciation in nature (Nosil, 2012). The process is based on the premise that traits evolving repeatedly and independently in multiple descendant populations that inhabit similar habitats (Schluter & Nagel, 1995) also reproductively isolate the descendant populations from their ancestral population. Therefore, during parallel ecological speciation populations inhabiting similar environments are expected to be reproductively compatible, whereas those inhabiting dissimilar environments are expected to be reproductively isolated (Schluter & Nagel, 1995). Ecological parallel speciation distinguishes from other types of parallel speciation in that no other force (e.g. genetic drift) could have favoured the evolution of RI between ancestor and derived species in a replicate manner.

In a recent review, Ostevik and colleagues (2012) outlined the standard of evidence required by any system to be considered an example of parallel ecological speciation, and highlighted how such evidence is lacking in plants. Phylogenetic independence of each diverging pair from other pairs, or in other words, evidence that populations that inhabit similar environments do not form a monophyletic group, is perhaps the most common result described for most plants. However, not all phylogenetic studies have used multilocus data (Nicholls & McNeilly, 1982; Pérez, 2011; Ostevik

et al., 2012), and in many cases it is difficult to rule out paraphyly due to gene flow instead of being caused by independent origins. A more difficult line of evidence comes from the level of intrinsic RI between independently evolved populations inhabiting divergent habitats. Such genetically based RI between descendent and ancestral populations is in most cases difficult to measure as many plants cannot be easily grown and crossed over generations in controlled conditions. Further, the cases where intrinsic RI barriers have been measured might not correspond to cases of parallel evolution, thus creating a natural decoupling between genetic and phylogenetic studies. Some studies have looked at other forms of RI, particularly to those arising as a consequence of local adaptation (e.g., immigrant inviability and/or F1 extrinsic postzygotic isolation), but such results are often considered weak evidence given that extrinsic RI barriers are considered labile and reversible. The other side of the coin comes from lack or low levels of intrinsic and extrinsic RI between descendent populations, or those inhabiting similar environments. This is a key criterion that distinguishes parallel ecological speciation from other forms of parallel speciation, such as mutation-order speciation (Schluter, 2009).). Finally, cases of parallel ecological speciation face the great difficulty in identifying the trait that evolves under natural selection and that simultaneously creates RI between ancestral and descendant populations. Usually, this data is explored in reciprocal transplants where traits underlying local adaptation are identified and linked to causes of extrinsic RI. More difficult, though, is to link such traits to the evolution of intrinsic RI, as this will likely require genetic studies that test for co-localisation, or linkage, between those gene causing extrinsic and intrinsic RI, and evidence for molecular signature of selections in the genes underlying variation in such traits (Nosil, 2012). Although in some cases of animal speciation these criteria (i.e., phylogenetic independence, isolation, compatibility, and selection) have been fulfilled, in plants they seem to be scarce (Ostevik *et al.*, 2012), possibly suggesting that speciation differs between plants and animals, or that the ecological speciation hypothesis has not been formally tested in plants. In this chapter we directly test for two of the criteria, isolation and compatibility in the *Senecio lautus* system, and use data from my previous chapter to argue that we have found a strong case of parallel ecological speciation in an herbaceous plant.

Senecio lautus is a young complex of groundsels inhabiting a wide variety of environments in Australia (Ali, 1966; Ali, 1968; Ali, 1969; Radford *et al.*, 2004; Thompson, 2005). Within the complex, a system of populations inhabiting multiple Dune and Headland environments exhibit some key attributes proper of systems undergoing parallel speciation. Populations display morphological similarity according to the environment in which they are found, thus suggesting a correlation between habitat and trait variation. However, such correlation is found independently of phylogenetic relations (Roda *et al.,* 2013a and Fig. 3.1 from previous chapter) suggesting that traits

like growth habit have evolved repeatedly and independently multiple times in the system. As shown in Chapter II, and in similar reciprocal transplant experimental results from the Ortiz-Barrientos Lab (Walter and Ortiz-Barrientos unpublished results), we have detected extrinsic reproductive barriers preventing gene flow between Dune and Headland populations. Furthermore, we have found that herbivores and possibly edaphic conditions create both immigrant inviability and selection against hybrids (Melo *et al.*, 2014), which together suggest weak forms of isolation and traits conferring RI between ancestral and derived populations in the *S. lautus* system. Below, we present the results from an experiment designed to formally test for RI and compatibility between ancestral and derived populations that are considered independent instances of evolution and that occupy similar and contrasting habitats along the eastern and southern coasts of Australia.

MATERIALS AND METHODS

The system and the test

Dune and Headland populations of *Senecio lautus* show strong leaf and plant architecture correlation with the different environments (Radford *et al.*, 2004). These morphological differences are genetically based and are preserved when populations are grown under glasshouse conditions. Dune adapted populations are tall, erect or decumbent, and poorly branched. Populations adapted to the Headlands are short, prostrate and heavily branched. Dune habitats are characterized by sandy soil poor in nutrients, which easily drains water and reaches high temperatures due to sunlight exposure. Headland habitats are subject to strong winds, constantly sprayed by salty water and its soil contains many nutrients.

Similar selective pressures resulting from unique environmental similarities among Dune or among Headland habitats (Roda *et al.* 2013a) could be responsible for the replicated evolution of these morphs. Previous phylogenetic and population structure analysis indicated that genetic variability structures into two main clades that follow the geographic distribution of the eastern and southern coasts of Australia, respectively. Although Dune and Headland populations from the eastern clade are genetically compatible in the glasshouse, they display strong extrinsic RI in their natural habitats (selection, Melo *et al.*, 2014; Walter et al. unpublished data). However, we remain ignorant as to whether populations from the southern coast follow similar reproductive patterns, and whether crosses between individuals from the two major independent clades are reproductively compatible with each other. Here we use the coastal ecotypic system in *Senecio lautus* to answer these questions, but more specifically to test whether natural selection is playing a role on the overall

speciation process in this system. Below we describe the experimental populations and the crosses performed here as well as the measurements of RI we took.

We used individuals of eight populations, four of the eastern coast (D01-H01 and D04-H05) and four of the southern coast (D32-H12 and D23-H21) of Australia (see Table 4.1 for geographic locations) to perform the following test: If natural selection is driving the evolution of RI in response to adaptation to contrasting habitats, we expect that populations adapted to similar environments will be reproductively compatible, but those adapted to contrasting habitats will be reproductively isolated. We expect to see this trend both in crosses within and in crosses between clades (Fig. 4.1a), where we have the strongest evidence for phylogenetic independence for the origin of Dune and Headland populations (Chapter III and Roda *et al.* 2013a,b). Because previously reported data (Melo *et al.*, 2014) showed that crosses within and between ecotypes were compatible between individuals derived from the same eastern clade, we expect certain modifications to the traditional predictions of the ecological speciation hypothesis (Fig. 4.1b). First, pattern of reproductive compatibility and isolation might be similar between southern and eastern clades. Phylogenetic relationships and estimated divergence times suggests that Dune and Headland populations have followed similar trajectories. Furthermore, Dune and Headland habitats are less contrasting in the southern than in the eastern coast (Roda *et al.*, 2013a). However, we still expect reproductive compatibility within ecotype but not between ecotypes in crosses of individuals derived from different clades (Fig. 4.1b) if parallel speciation is driven by divergent natural selection. We evaluated reproductive compatibility and isolation by measuring F1 seed set, immigrant inviability, and F1 seed viability. We did not measure other forms of F1 dysfunctions so our results should be interpreted within the realm of the traits explored here.

Experimental populations

Seeds from 30 individuals were collected for the eight populations in the study directly at their respective natural environments (each individual is considered a family). Seeds were scarified (1mm trimmed at the micropyle side) and germinated on moist filter paper in petri dishes. Seeds were kept in dark and controlled conditions for three days to induce root elongation, and subsequently placed under the light for 7 days to induce vegetative growth. One week old seedlings were transferred to a glasshouse with constant temperature (25°C) and 12h:12h light:dark cycle and transplanted into 0.25 L pots filled with standard potting mix. After two months, flowered individuals were crossed to create experimental seed stocks for each population –having removed the effect that maternal provisioning could have on individual's performance. Previous reports have found that coastal ecotypes exhibit strong self-incompatibility (Ornduff, 1964; Melo *et al.*, 2014).

Reproductive isolating barriers

To investigate patterns of RI in the *Senecio lautus* complex, we chose three different reproductive isolating barriers to study, including both intrinsic *(I)* and extrinsic *(E)* reproductive barriers: 1) immigrant inviability *(E)*, 2) F1 hybrid seed set *(I)* and 3) F1 hybrids inviability *(I)*. These measurements were taken within and between clades, and from the basis of the test explained above. This level of comparison relies on phylogenetic independence and ensures that between clade comparisons are of similar age. Overall, we tested for the effect of ecotype (different or same ecotype), region (within and between region) and divergence time (this last as a covariate) on the strength of RI of each barrier applying this linear model using JMP 10.0.2 (SAS Institute Inc.)

Immigrant inviability due to soil differences

Because soil is one of the most notable differences between Dune and Headland habitats (Roda *et al.*, 2013b) –and its effect is easy to assess independently from other ecological variables– here we investigated its effects on the germination of eight coastal populations (four parapatric pairs) of *Senecio lautus*. We performed a transplant experiment following a full factorial design, evaluating the ability of all seeds of all populations to germinate on the soil from the eight localities (soil from four Dune habitats and four Headland habitats). Experiments were run under controlled conditions (25°C through a 12h: 12h light:dark cycle) in the University of Queensland (QLD, Australia), filling eight plastic trays (30 wells each), each with soil from the different localities (Table 4.1). The soil was pasteurized (60 C for 76 hours); to eliminate pathogens that could have arisen during the time it was stored before it was used. Seeds belonging to 25 families of each D01, D04, D23, H01, H05 and H21; 7 of D32 and 18 of H12 were sown into each of the trays (each family was represented by one seed per tray) for a total of 175 seeds per tray and 1400 in the all trays. Seeds were sown on top of the soil in a fully randomized design within each tray. Trays were sprayed daily to keep the soil moist, and tray position on the shelf was also switched every day. Seed germination occurred during the first 66 days of the experiment.

F1 hybrid seed set

15 families for each of the populations were germinated and grown under glasshouse conditions (described above). Intra and interpopulation crosses (Table S4.1) were performed twice a day by gently rubbing flower heads (capitula): each flower head was crossed at least three times to saturate the number of fertilized florets. Seeds (empty and filled, Fig S2.2) where stored in coin envelopes for subsequent counting. We calculated seed set by estimating the proportion of fertilized seeds in flower heads (Fig. S2.2). We divided the number of fertilized seeds in an interpopulation cross by

the average number of seeds produced in parental intrapopulation crosses; we subtracted this fraction from one to calculate postmating prezygotic isolation in the system (Coyne & Orr, 2004).

F1 viability at early stages of the life cycle

We tested for the viability of F1 hybrids between multiple interpopulation crosses (see above), counting the proportion of seeds that germinated in relation to the parental germination proportions. A total of 260 families of all cross types were germinated placing five seeds of each family into moist filter paper in petri dishes (one family per petri dish). Then we placed petri dishes into trays (20 per tray) randomizing the position of all families. Germination conditions used here are the same as described above, except for tray position on the shelves, which in this case was switched daily. Also we did not scarified seeds to truly investigate on the intrinsic ability of embryos to germinate. We calculated F1s inviability by estimating the proportion of germinated seeds in a cross, divided by the average number of seeds that germinated of the two parental populations.

Strength of RI

We calculated the strength the three RI barriers amongst coastal populations following the approach by Lowry *et al.*, (2008). Estimates were done for the three reproductive isolating barriers in the following way: *Immigrant inviability (E)*, or the ability of migrant seeds to germinate in the alternative population soil, as $RI_{imm}=1-(w_i/w_n)$, where w_i is the mean number of migrant individuals that germinated, and w_n the of the local population. *Hybrid seed set (I)* in the glasshouse or whether the proportion of fertilized seeds in a flower head from an inter-population cross differed from an intra-population cross as $RI_{seed\ set}=1-(Pf_{inter}/Pf_{intra})$ were Pf stands for proportion fertilized. *Hybrid viability (I)*, or whether hybrid seedlings germinated equally well as their parents in the glasshouse, as $H_{hf}=1-(v_{F1}/v_{parents})$, where v_{F1} is the average germination of F1 hybrids, and $v_{parents}$ is the average germination of the two parents. We also calculated the cumulative total RI of the three barriers in the study. For this we estimated the relative strength of the three barriers according to the order in which they occur in a population's life cycle (Ramsey *et al.*, 2003). In this case the first barrier to arise is immigrant inviability, followed by F1 seed set and the last is F1 inviability.

RESULTS

Immigrant Inviability

For all population comparisons, the average strength of RI for immigrant inviability corresponded to negative values, reflecting a higher germination of the immigrant population than for the local (Fig. 4.2, Table 4.2). We found a pattern of increasing negative values for *comparisons between regions* to *within regions*, and for *different ecotype* to *same ecotype* (Fig. 4.2, Table 4.2), but neither ecotype (same/different) nor region (within/between) had a significant effect on RI of this barrier (Table 4.3). When we separated groups depending on whether the migrant population was a Dune or a Headland population, the only comparison that did not follow the negative RI trend were Dune individuals germinating at Headland habitats *between regions*, which exhibited an average positive value of RI (1.0 +/- 0.092, Table 4.2).

F1 hybrid seed set

We found that crosses *within region* presented on average negative values of RI, with more negative values for crosses of the *same ecotype* (Fig. 4.3, Table 4.2). This indicates that populations *within region* produced a higher proportion of fertilized seeds than the pure parental crosses, particularly when they came from Dune x Dune and Headland x Headland crosses. Contrastingly, crosses *between regions* on average produced less fertilized seeds than the pure parental crosses – suggested by the positive values of RI– and in particular when crosses were between ecotypes (Dune - Headland, 0.906 +/- 0.03, Table 4.2). We found a significant effect of ecotype and divergence time on the RI strength for this barrier (ecotype: estimate= 0.38, $t = 4.39$, $P < 0.0001$; divergence time: estimate = 0.000004, $t = 2.41$, $P = 0.021$). The positive estimates of RI for population crosses from the *same ecotype* and *between regions* were driven by Headland x Headland crosses (0.519 +/- 0.046), and not between Dunes (-0.207 +/- 0.182, Table 4.2). Again, ecotype had a significant effect on the RI for crosses with Dune mother and marginally significant for the ones with a Headland mother (Dune mother: estimate = 0.465, $t = 4.51$, $P < 0.001$; Headland mother: estimate = 0.255, $t = 1.93$, $P = 0.074$), while divergence time had a significant effect only when crosses had a Dune mother (Table 4.3).

F1 hybrid inviability

In the case of F1 viability, all comparison types exhibited on average positive values of RI, indicating that pure parental crosses produce more viable seeds than hybrid ones (Fig. 4.4, Table 4.2). Again the highest estimates of RI were for crosses from *different ecotypes* and *between regions*, while other crosses had variable values of RI (Fig. 4.4, Table 4.2). There was a significant interaction between ecotype and region on the strength of F1 inviability (estimate: 0.185, $t = 3.16$, P

= 0.004). Interestingly, the hybrid product of crosses between Dune populations had on average higher proportion of viable seeds (Dune x Dune: between regions -0.166 +/- 0.152, within regions (-0.064 +/- 0.071) than crosses between headlands (Headland x Headland: between region: 0.139 +/- 0.07, within region 0.187 +/- 0.221, Table 4.2). Crosses with a Dune mother also had a significant interaction between ecotype and region (estimate = 0.207, t = 2.52, P = 0.027, Table 4.3).

Total cumulative RI

Taking into account the three RI barriers in a total cumulative estimate of RI, comparisons of populations *between regions* exhibited positive estimates of RI, while populations *within region* were on the negative scale (Fig. 4.5, Table 4.2). The highest estimates for RI were for population comparisons from *different ecotypes* and *between regions* (0.941 +/- 0.261, Table 4.4) and the most negative estimates for RI were for comparisons of the *same ecotype* and *within region* (-0.493 +/- 0.394, Table 4.4). Ecotype had a significant effect on the total RI (estimate = 0.296, t = 2.36, P = 0.022, Table 4.3).

See Table S4.2 for estimates of RI for each of the population comparisons for each reproductive barrier and for the cumulative.

DISCUSSION

Parallel evolution of RI is one of the strongest lines of evidence for the role of natural selection in speciation (Schluter & Nagel, 1995). Although arguing for parallel ecological speciation requires evidence for many experimental fronts (Ostevik *et al.*, 2012), the process seems to be possible, and perhaps is common in animals (Rundle *et al.*, 2000; Nosil *et al.*, 2002; Richmond & Reeder, 2002; Boughman *et al.*, 2005; Johannesson *et al.*, 2010; Ostevik *et al.*, 2012). This is not the case in plants where cases with strong evidence for parallel ecological speciation seam rather rare (but see Foster *et al.*, 2007). Previous studies in *Senecio lautus* suggested that coastal and alpine populations could be evolving parallel ecotypes in several localities of Australia (Roda *et al.*, 2013a; Melo *et al.*, 2014). Here, we provide evidence that natural selection is driving the parallel evolution of RI between Dune and Headland ecotypes in Australia. In our test of parallel ecological speciation, we found that populations that inhabit similar environments are more reproductively compatible, while populations inhabiting different environments exhibit more RI, especially when comparisons included populations from different geographic regions. We found that the strength of RI for ecotypic comparisons is consistent with the prediction for parallel ecological divergence within

region and with parallel ecological speciation between regions. We suggest that *S. lautus* is a good candidate for the parallel ecological speciation and an excellent model to study speciation with gene flow.

Variable patterns of RI in Senecio lautus

We found that *S. lautus* coastal ecotypes from different regions or from the same region showed variable patterns of RI. Notably, Dune and Headland seeds germinated similarly in soil derived from sand dunes or rocky headlands regardless of locale of origin (Fig. 4.2). However, production of hybrid seed or viability of hybrid seed was dramatically reduced in crosses between different ecotypes from different regions (Fig 4.3. and Fig. 4.4). In contrast, crosses between individuals of the same ecotype but from different regions produced similar levels of F1 seed set and viability to those produced by within population crosses (Fig. 4.3 and Fig. 4.4). Previous studies in this system are largely consistent with these results: intrinsic RI was weak within the eastern region, and soil had little effect on germination in transplant experiments under both field and glasshouse conditions. In these previous experiments, RI between Dune and Headland populations was strong but extrinsic, and occurred at later stages of divergence after seed germination. (Melo *et al.* 2014; Walter and Ortiz-Barrientos unpublished results).

Complex variable patterns of RI are explained in straightforward fashion using the logic of ecological speciation. First, greater RI between regions and ecotypes is expected under both the deep phylogenetic divergence between clades and the differential adaptation to sand dunes and rocky headlands. Similarly, lack of RI between individuals of the same ecotype, and regardless of phylogenetic divergence, is expected under parallel adaptation of such populations (Schluter & Nagel, 1995; Ostevik *et al.*, 2012). It is remarkable, however, that such compatibility has persisted over several hundred thousand years. Although seemingly unexpected, lack of RI within region and regardless of cross type can be expected if stages of speciation are variable across the system: young divergences are taking place within regions, but older divergences continue to accumulate between regions. Reciprocal transplant experiments within regions are consistent with this view, and suggest that extrinsic RI may characterise the first stages of speciation whereas intrinsic RI the later stages of speciation.

Extrinsic first, intrinsic second

Lack of intrinsic RI within regions suggests that local adaptation my not directly lead to the evolution of intrinsic RI. This is not predicted by the ecological speciation hypothesis where divergent natural selection leads to the evolution of reproductive incompatibilities between

populations adapting to contrasting environments. It is possible that such incompatibilities are segregating at later stages of hybridisations, such as the F2 generation, and thus we did not sample them in our experiment, which would not be surprising given that hybrid breakdown (beginning in the F2 generation) is common in plants (Mackill & Ni, 2001; Johansen-Morris & Latta, 2006). Alternatively, intrinsic RI evolves because of different reasons.

Intrinsic RI could evolve only when gene flow is very limited or lacking between populations, a firmly supported theoretical result (Thibert Plante & Hendry, 2010). Lack of gene flow in parapatry can evolve in two ways. First, it can evolve in a localised way in response to the effects of selection around a locus. Because selection prevents gene flow outside the locus of interest due to linkage, other mutations that would not normally persist under gene flow can now diverge between populations. Thus, it is possible that mutations contributing to intrinsic RI evolved in tight linkage with those under selection. A problem with this scenario to explain our results is that although lack of intrinsic RI within clades is an acceptable option over certain periods of time, it is very hard to explain how deep divergences would remain compatible under this scenario. For this to work, the neighbourhood of genes around the selected site, as well as the mutational availability, would have to be highly constrained within a given habitat so crosses between two Dune or two Headland populations from different regions (and clades) remained reproductively compatible.

Intrinsic RI could also evolve from general reductions in gene flow across the entire genome. Genomic reductions in gene flow can arise from strong divergent natural selection on phenotypes, or from genome hitchhiking. In either case (instantaneous genomic isolation or rapid whole genome isolation through the accumulation of many selective mutations) genetic drift will be free to drive the evolution of many new mutations anywhere in the genome of parapatric populations (Feder *et al.*, 2012). Therefore, it would not be surprising to evolve intrinsic RI after a period of isolation, even if in parapatric conditions. This scenario is more problematic for the ecological speciation hypothesis, because patterns of parallel speciation would be harder to explain as mutational input between populations adapted to the same environment would be different and thus these populations would be expected to become reproductively isolated as well.

So, how can we explain strong isolation between ecotypes only when they are derived from different regions, and lack of intrinsic RI in crosses between individuals of the same ecotype regardless of where they come from? Surprisingly, we might need to slightly modify the traditional Dobzhansky-Muller model (Fig. 4.6a) for the evolution of intrinsic RI. The simpler version of the DM model states that an ancestor carrying the epistatic genes *AABB* might lead to descendant

populations each with a new fixed allelic version of the epistatic genes. In one population the genotype becomes *AAbb*, whereas in the second population it becomes *aaBB*. Intermediate stages (i.e., *AaBB* and *AABb*) are at least as fit as the ancestral genotype, and only the fully hybrid genotype, *AaBb*, is unfit, thus leading to intrinsic RI between the two descendant populations (Fig. 4.6a). This simple model cannot account for our results, as it would lead to intrinsic RI between ecotypes both within and between regions. We propose that a system with two pairs of epistatic genes, *AABB* and *CCDD* (Fig. 4.6b), are sufficient for creating the patterns we see in the *Senecio lautus* system (Fig. 4.3).

Consider only one Dune and Headland pair in each region. Mutations in the epistatic genes can either occur before the split between the two regional clades, within a clade but before the split of Dune and Headland populations, or in a Dune or Headland population. Mutations in the first two scenarios will lead to intrinsic RI between ecotypes both within and between regions, so they can be refuted. However, mutations leading to the evolution of Dune and Headland populations lead to sensible models. Specifically, we only require that Dune and Headland populations from the different clades experience mutations on the same epistatic gene pair (Fig. 4.6b). For instance, the Dune from clade one and the Headland from clade two will experience substitution in the AABB system, whereas the Headland from clade two and the Dune from clade one will experience substitutions in the second epistatic gene pair (CCDD). Thus, crosses between ecotype within region will be reproductively compatible, but crosses between ecotypes between regions will be reproductively isolated. More importantly, crosses between the same ecotype will be reproductively compatible regardless of region (Fig. 4.6).

Criteria for parallel ecological speciation

Experiments shown here and in previous reports (including chapters in this dissertation) suggest that the *S. lautus* complex is a case of parallel ecological speciation. Classic work by Schluter and Nagel (1995), and recent reviews on ecological speciation (Nosil, 2012; Ostevik *et al.*, 2012) have delineated clearly the conditions for claiming parallel ecological speciation in a system:

Independence is probably the most common line of evidence that both animals and plant systems meet. It indicates that traits isolating populations are evolving independently and are not inherited from a common ancestor. Although simple phylogenetic analyses provide evidence for it, strong evidence for independence should arise from robust phylogenetic studies that include multiple loci across the genome (Ostevik *et al.*, 2012). This is especially important in systems that have undergone recent divergences and/or experienced high levels of hybridization. However, various

candidates of parallel ecological speciation in plants frequently present weak evidence of independence often presenting evidence based on few chloroplast markers (Vijverberg *et al.*, 1999; Noguchi & De-yuan, 2004; Pérez, 2011). Markers like this are inherited as a single locus and could be acquired by horizontal gene transfer (Stegemann *et al.*, 2012). Other plant systems like *Eucalipthus globus, Lasthenia californica and Armeria maritime* successfully fulfil this condition (Rajakaruna & Whitton, 2004; Baumbach & Hellwig, 2007; Foster *et al.*, 2007), as well as various cases in animal systems (Schluter & Nagel, 1995; Colosimo *et al.*, 2005; Quesada *et al.*, 2007; Ostevik *et al.*, 2012; Strecker *et al.*, 2012).

Consistent with the concept of ecological speciation, where RI evolves between populations adapting to contrasting environments, parallel ecological speciation predicts that in systems showing parallel evolution of ecotypes, RI should be higher amongst populations at contrasting habitats than at similar habitats (Schluter & Nagel, 1995). To provide strong evidence for isolation, studies should be able to detect both intrinsic and extrinsic RI, and if only one of them is found the evidence is weak. So far we found that most studied plant systems to date lack strong evidence for isolation, this could be a consequence of: first, very few plant systems have directly aimed to study RI (Sobel *et al.*, 2010), therefore systems that exhibit repeated morphologies at multiple localities might not have been tested for either intrinsic or extrinsic RI. Secondly, intrinsic postzygotic isolation may be difficult to detect, as there is a strong possibility that it simply has not evolved yet in systems with recent divergences. In 2012, Levin provided an overview of the time it takes for intrinsic RI to evolve in flowering plants. He explained that it sometimes takes millions of years (>4 million years) after divergence (although intrinsic hybrid inviabiliy rapidly evolved between tolerant and non-tolerant populations in *Mimulus*, Macnair & Christie, 1983). Lastly, evidence for RI could be mitigated by hybrid vigour (Lowry *et al.*, 2008) a common phenomenon in plants (Raabová *et al.*, 2009). Overall, plant systems that present evidence for selection is weak, either reporting intrinsic RI (Pérez, 2011) or extrinsic RI (Vijverberg *et al.*, 2000; Foster *et al.*, 2007). In invertebrates, *Littorina saxatilis* exhibits both intrinsic and extrinsic reproductive barriers (Rogers & Bernatchez, 2006), but in animals it is more common to find only extrinsic RI (Rundle *et al.*, 2000; Nosil *et al.*, 2002).

Intimately related to the previous condition, under parallel ecological speciation populations of the same ecotype are expected to be more reproductively compatible, or less reproductively isolated than populations of different ecotypes. Despite being particularly informative, evidence for compatibility is scarcer. If a candidate system fails to find reproductive compatibility between

daughter populations of the same ecotype, this could mean that unique mechanisms of isolation are driving speciation in each diverging pair, opposed to natural selection driving it (Schluter & Nagel, 1995). However, the likelihood of detecting compatibility depends on the stage of speciation of the system. For instance, populations that inhabit allopatric yet similar environments but are at advanced stages of speciation will eventually accumulate random genetic differences and become isolated. Thus, it is possible that the evolution of intrinsic RI reduces the likelihood of detecting compatible populations found in similar habitats. Similarly to isolation, a lack of studies of RI in candidate systems in plants could contribute to the lack of evidence for compatibility. However, systems like *Lasthenia californica* exhibit low seed set amongst populations of different ecotypes and higher pollination success amongst populations of the same ecotype. Although this is considered weak evidence for both isolation and compatibility, this is probably the best candidate for parallel ecological speciation in plants (Rajakaruna & Whitton, 2004). Compatibility is also the least supported condition in animals (Rundle *et al.*, 2000; Nosil *et al.*, 2002; Rolán-Alvarez *et al.*, 2004).

Finally, to gain the evidence required for parallel ecological speciation, studies should identify an adaptive mechanism and test for it. This is direct evidence demonstrating the specific role of ecology in speciation, but it is also useful to distinguish parallel ecological speciation from other cases of parallel evolution that involve mechanisms such as polyploidy (Schluter & Nagel, 1995). Although most of the evidence for selection in plants arises from the correlation between morphologies and environmental conditions, a few systems have provided strong evidence for the selection criterion, having even found the loci responsible for the adaptive trait (e.g. the two loci responsible for copper tolerance in Silene vulgaris, Schat *et al.*, 1993). Similarly, studies in animals have also been successful at the same level (Colosimo *et al.*, 2005; Chan *et al.*, 2010; Renaut *et al.*, 2011; Rogers *et al.*, 2013).

Overall, most of the plant systems reviewed by Ostevik *et al.* (2012) for parallel ecological speciation in plants had only weak or indirect evidence for any of the criteria. For the first time in a plant system, *Senecio lautus* presents strong evidence for all the criteria, constituting the best-known example for parallel ecological speciation in plants. I summarise the main arguments for this: 1) The signature for the multiple independent origins in the coastal system stems on separate and robust studies using multilocus data from thousands of RAD markers across genomes and 13 neutral markers. This makes independence a strong argument in *S. lautus.* 2) *S. lautus* exhibits strong evidence for isolation from both extrinsic and intrinsic RI. For extrinsic RI we have found

that various parapatric pairs presented strong reproductive barriers in the field, as reported in a previous chapter (Melo *et al.*, 2014). For intrinsic RI specifically for F1 seed set, we found that the type of cross (different ecotype or the same ecotype) strongly influenced the strength of the barrier. However it was only in crosses of populations between the two geographic regions where we found evidence for isolation (Fig. 4.3). Although this pattern was not found in crosses within the same region, the statistical test did not detect an effect of region (Table 4.3), as the pattern was similar but on the negative scale (Fig. 4.3, see below for compatibility evidence). Hybrid inviability followed the same trend in comparisons between regions (Fig. 4.4). 3) *Senecio lautus* offers extensive evidence for reproductive compatibility. We found that for the F1 hybrid seed set, the same ecotype crosses were found more compatible than different ecotype crosses, both with and between regions (Fig. 4.2 – Fig. 4.5). However, we observed that interpopulation crosses within the same region would sometimes produce even more and healthier seeds than pure crosses. These results suggest that populations from the same region are highly reproductively compatible as previously reported (Melo *et al.*, 2014). We propose that in young divergences like in *S. lautus,* the compatibility criterion is more important to the model of parallel ecological, specially considering that intrinsic RI may take a long wait to evolve, as it has been suggested for plants (Levin, 2012), and for other groups (Mallet, 2006). For example, it can be observed for some of the barriers studied in here, where comparisons are compatible but same ecotypes are more compatible than different ecotypes. Although hybrid vigour appears to be common in the *S. lautus* complex (Melo *et al.*, 2014), it is possible that it is only apparent, and that hybrids at later stages could result inviable (or infertile). 4) Finally, selection in the system is supported by mixed lines of evidence: First, in transplant experiments in the field where Dune, Headland and Hybrid seed were sowed in both environments, we found that predation killed migrant and hybrid seedlings in a higher proportion compared to the local (Melo *et al.*, 2014). Other studies in our lab have also suggested that edaphic selection – specially related to substrate salt content– could be related to the differential performance of ecotypes in the alternative environments and to differentiated genetic variation in the ecotypes (Fig. 2.1., Melo et al. in Press; Roda *et al.*, 2013b; Brittain et al. unpublished). Although both mechanisms are to be studied in more detail, this and other environmental factors could be involved in the local adaptation of plant populations.

Coastal populations in *Senecio lautus* constitute an excellent case of parallel ecological speciation in plants. Our results suggest that each Dune and Headland pair is evolving independently and by the action of natural selection. We found that both ecology and geographic clade have an effect on the evolution of RI. It is likely that more intrinsic RI evolves between populations separated by older divergences than between populations that have recently diverged (if we note that Dune and

Headland parapatric populations have stopped exchanging genes). Further, we suggest that a system with two pairs of epistatic genes, *AABB* and *CCDD*, are sufficient for creating the patterns we see in the *Senecio lautus* system. These particular findings suggest that intrinsic reproductive isolation may be decoupled from the process of speciation, arising long after populations have diverged. Overall our results indicate that extrinsic barriers play a protagonist role in speciation with gene flow. Finally, we suggest that ecological speciation may be more similar between animals and plants than previously thought.

Acknowledgements

This research was funded by the Australian Research Council grants to D.O-B. Thomas Richards and Greg Walter collected seeds from the natural populations and soil samples across the Australian coasts. We thank the glasshouse services unit at the University of Queensland for assistance with pest control and glasshouse space maintenance throughout several stages of the experiments.

References

Agrawal AF, Feder JL, Nosil P. 2011. Ecological divergence and the origins of intrinsic postmating isolation with gene flow. *International Journal of Ecology* **2011**.

Ali S. 1966. *Senecio lautus* complex in Australia. III. The genetic system. *Australian Journal of Botany* **14**: 317-327.

Ali S. 1968. *Senecio lautus* complex in Australia. IV. The biology of the complex. *Phyton (Horn, Austria)* **13**: 53-62.

Ali SI. 1969. *Senecio lautus* complex in Australia. V. Taxonomic interpretations. *Australian Journal of Botany* **17**: 161-176.

Baumbach H, Hellwig F. 2007. Genetic differentiation of metallicolous and non-metallicolous *Armeria maritima* (Mill.) Willd. taxa (Plumbaginaceae) in Central Europe. *Plant Systematics and Evolution* **269**: 245-258.

Boughman JW, Rundle HD, Schluter D. 2005. Parallel evolution of sexual isolation in sticklebacks. *Evolution* **59**: 361-373.

Chan Y, Marks M, Jones F, Villarreal Jr G, Shapiro M, Brady S, Southwick A, Absher D, Grimwood J, Schmutz J. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**: 302.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**: 1928-1933.

Coyne J, Orr H. 1998. The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **353**: 287.

Coyne JA, Orr HA. 2004. *Speciation*: Sinauer Associates Sunderland, MA.

Feder JL, Gejji R, Yeaman S, Nosil P. 2012. Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 461-474.

Foster SA, McKinnon GE, Steane DA, Potts BM, Vaillancourt RE. 2007. Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. *New Phytologist* **175**: 370-380.

Johannesson K, Panova M, Kemppainen P, André C, Rolán-Alvarez E, Butlin RK. 2010. Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1735-1747.

Johansen-Morris AD, Latta RG. 2006. Fitness consequences of hybridization between ecotypes of *Avena barbata:* hybrid breakdown, hybrid vigour, and transgressive segregation. . *Evolution* **60**: 1585-1595.

Levin DA. 2012. The long wait for hybrid sterility in flowering plants. *New Phytologist* **196**: 666-670.

Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH. 2008a. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 3009-3021.

Mackill D, Ni J. 2001. Molecular mapping and marker-assisted selection for major-gene traits in rice. *Rice genetics IV*: 131-151.

Macnair M, Christie P. 1983. Reproductive isolation as a pleiotropic effect of copper tolerance in Mimulus guttatus. *Heredity* **50**: 295-302.

Mallet J. 2006. What does Drosophila genetics tell us about speciation? *Trends in Ecology & Evolution* **21**: 386-393.

Melo MC, Grealy A, Brittain B, Walter GM, Ortiz-Barrientos D. 2014. Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsell. *New Phytologist* **doi: 10.1111/nph.12779**.

Nicholls MK, McNeilly T. 1982. The possible polyphyletic origin of copper tolerance in *Agrostis tenuis* (Gramineae). *Plant Systematics and Evolution* **140**: 109-117.

Noguchi J, De-yuan H. 2004. Multiple origins of the Japanese nocturnal *Hemerocallis citrina* var. vespertina (Asparagales: Hemerocallidaceae): evidence from noncoding chloroplast DNA sequences and morphology. *International Journal of Plant Sciences* **165**: 219-230.

Nosil P. 2008. Speciation with gene flow could be common. *Molecular Ecology* **17**: 2103-2106.

Nosil P. 2012. *Ecological speciation*: Oxford University Press.

Nosil P, Crespi BJ, Sandoval CP. 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* **417**: 440-443.

Ornduff R. 1964. Evolutionary pathways of the *Senecio lautus* alliance in New Zeland and Australia. *Evolution*: 349-360.

Ostevik KL, Moyers BT, Owens GL, Rieseberg LH. 2012. Parallel ecological speciation in plants? *International Journal of Ecology* **2012**.

Pérez F. 2011. Discordant patterns of morphological and genetic divergence in the closely related species *Schizanthus hookeri* and *S. grahamii* (Solanaceae). *Plant Systematics and Evolution* **293**: 197-205.

Quesada H, Posada D, Caballero A, Morán P, Rolán-Alvarez E. 2007. Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* **61**: 1600-1612.

Raabová J, Münzbergová Z, Fischer M. 2009. Consequences of near and far between-population crosses for offspring fitness in a rare herb. *Plant Biology* **11**: 829-836.

Radford I, Cousens R, Michael P. 2004. Morphological and genetic variation in the *Senecio pinnatifolius* complex: are variants worthy of taxonomic recognition? *Australian Systematic Botany* **17**: 29-48.

Rajakaruna N, Whitton J. 2004. Trends in the evolution of edaphic specialists with an example of parallel evolution in the *Lasthenia californica* complex. *Plant adaptation: Molecular biology and ecology*: 103-110.

Ramsey J, Bradshaw Jr H, Schemske D. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**: 1520-1534.

Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L. 2011. SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (Coregonus spp.). *Molecular Ecology* **20**: 545-559.

Richmond JQ, Reeder TW. 2002. Evidence for parallel ecological speciation in scincid lizards of the *Eumeces skiltonianus* species group (Squamata: Scincidae). *Evolution* **56**: 1498-1513.

Roda F, Ambrose L, Walter GM, Liu HL, Schaul A, Lowe A, Pelser PB, Prentis P, Rieseberg LH, Ortiz-Barrientos D. 2013a. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Molecular Ecology* **22**: 2941-2952.

Roda F, Liu H, Wilkinson MJ, Walter GM, James ME, Bernal DM, Melo MC, Lowe A, Rieseberg LH, Prentis P, Ortiz-Barrientos D. 2013b. Convergence and divergence during the adaptation to similar environments by an Australian groundsel. *Evolution* **67**: 2515-2529.

Rogers S, Bowles W, Mee J. 2013. The consequences of genomic architecture on ecological speciation in postglacial fishes. *Current Zoology* **59**: 53-71.

Rogers SM, Bernatchez L. 2006. The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (Coregonus clupeaformis). *Journal of Evolutionary Biology* **19**: 1979-1994.

Rolán-Alvarez E, Carballo M, Galindo J, Moran P, Fernandez B, Caballero A, Cruz R, Boulding EG, Johannesson K. 2004. Nonallopatric and parallel origin of local reproductive barriers between two snail ecotypes. *Molecular Ecology* **13**: 3415-3424.

Rundle HD, Nagel L, Boughman JW, Schluter D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**: 306.

SAS Institute Inc. JMP® 10.0.2. *Cary, NC, 1989-2007*.

Schat H, Kuiper E, Ten Bookum W, Vooijs R. 1993. A general model for the genetic control of copper tolerance in *Silene vulgaris*: evidence from crosses between plants from different tolerant populations. *Heredity* **70**: 142-142.

Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* **323**: 737.

Schluter D, Nagel LM. 1995. Parallel Speciation by Natural Selection. *The American Naturalist* **146**: 292-301.

Sobel J, Chen G, Watt L, Schemske D. 2010. The biology of speciation. *Evolution* **64**: 295-315.

Stegemann S, Keuthe M, Greiner S, Bock R. 2012. Horizontal transfer of chloroplast genomes between plant species. *Proceedings of the National Academy of Sciences*.

Strecker U, Hausdorf B, Wilkens H. 2012. Parallel speciation in *Astyanax* cave fish (Teleostei) in Northern Mexico. *Molecular Phylogenetics and Evolution* **62**: 62-70.

Thibert Plante X, Hendry A. 2010. When can ecological speciation be detected with neutral loci? *Molecular Ecology* **19**: 2301-2314.

Thompson I. 2005. Taxonomic studies of Australian *Senecio* (Asteraceae): 5. The *S. pinnatifolius/S. lautus* complex. *Muelleria* **21**: 23-76.

Via S. 2009. Natural selection in action during speciation. *Proceedings of the National Academy of Sciences* **106**: 9939-9946.

Vijverberg K, Kuperus P, Breeuwer J, Bachmann K. 2000. Incipient adaptive radiation of New Zealand and Australian Microseris (Asteraceae): an amplified fragment length polymorphism (AFLP) study. *Journal of Evolutionary Biology* **13**: 997-1008.

Vijverberg K, Mes TH, Bachmann K. 1999. Chloroplast DNA evidence for the evolution of *Microseris* (Asteraceae) in Australia and New Zealand after long-distance dispersal from western North America. *American Journal of Botany* **86**: 1448-1463.

Tables

Table 4.1. Populations of *Senecio lautus* in the study and their respective localities.

Locality	Coordinates	Population ID
	S 28° 48' 22.10" E 153° 36' 9.94"	H01
Lennox Head (NSW)	S 28° 47' 10.7" E 153° 35'	D01
	S 30° 18' 42.42" E 153° 8' 37.68"	H05
Coffs Harbour (NSW)	S 30° 18' 45.9" E 153° 08' 24.12"	D04
Portland, Cape Bridgewater (VIC)	S 38° 22' 49.6" E 141° 22' 07"	H12
Discovery Bay Costal Park (VIC)	S38° 19' 28.10" E141° 23' 42.80"	D32
	S 33° 09' 9.1" E 134° 15' 43.1"	H21
Point Labatt (SA)	S 33° 07' 30.9" E 134° 15' 57.0"	D23

Table 4.2. Strength for the three RI barriers in the study between coastal ecotypes of *Senecio lautus*. Estimates for immigrant inviability in the soil experiments only include seeds germination success. For F1 seed set and F1 inviability estimates were performed taking in account cross direction (Dune mother or Headland mother).

Barrier	Region	Ecotype	Comparison	N	Mean RI	S.E
Immigrant inviability	Between	Different	Dune migrant		0.1	0.092
			Headland migrant		-0.217	0.144
			All	16	-0.058	0.092
	Between	Same	Dune migrant		-0.045	0.17
			Headland migrant		-0.154	0.154
			All	16	-0.099	0.112
	Within	Different	Dune migrant		-0.039	0.164
			Headland migrant		-0.273	0.083
			All	16	-0.156	0.094
	Within	Same	Dune migrant		-0.089	0.144
			Headland migrant		-0.361	0.293
			All	8	-0.225	0.16
F1 seed set	Between	Different	Dune mother		0.873	0.076
			Headland mother		0.947	0.03
			All	9	0.906	0.044
	Between	Same	Dune mother		-0.207	0.182
			Headland mother		0.519	0.046
			All	14	0.104	0.143
	Within	Different	Dune mother		-0.023	0.297
			Headland mother		-0.106	0.305
			All	11	-0.061	0.203
	Within	Same	Dune mother		-0.234	0.079
			Headland mother		-0.485	0.3
			All	8	-0.359	0.151

Continuation Table 4.2. Strength for the three RI barriers in the study between coastal ecotypes of *Senecio lautus.* Estimates for immigrant inviability in the soil experiments only include seeds germination success. For F1 seed set and F1 inviability estimates were performed taking in account cross direction (Dune mother or Headland mother).

Barrier	Region	Ecotype	Comparison	N	Mean RI	S.E
F1 inviability	Between	Different	Dune mother		0.388	0.215
			Headland mother		0.439	0.038
			All	7	0.403	0.149
	Between	Same	Dune mother		-0.166	0.152
			Headland mother		0.139	0.07
			All	11	0	0.089
	Within	Different	Dune mother		0.07	0.093
			Headland mother		-0.026	0.043
			All	10	0.022	0.051
	Within	Same	Dune mother		-0.064	0.071
			Headland mother		0.187	0.221
			All	6	0.103	0.151

Table 4.3. Effects of region, ecotype, their interaction, and of divergence times on the RI for three reproductive barriers and the cumulative between population comparisons in *Senecio lautus*.

Barrier	Effects	Comparison	Estimate	T ratio	*P*
Immigrant inviability	Region	Dune mother	0.116	-0.69	0.496
		Headland mother	-0.038	-0.22	0.817
		All	-0.077	-0.64	0.524
	Ecotype	Dune mother	0.082	0.97	0.343
		Headland mother	0.027	0.32	0.754
		All	0.055	-0.68	0.371
	Ecotype*Region	Dune mother	-0.009	-0.11	0.913
		Headland mother	-0.059	0.68	0.503
		All	-0.034	-0.56	0.575
	Divergence Time	Dune mother	0.000	1.09	0.285
		Headland mother	0.000		0.501
		All	0.000	1.25	0.216
F1 seed set	Region	Dune mother	-0.327	-1.54	0.142
		Headland mother	0.332	1.16	0.266
		All	-0.053	-0.29	0.777
	Ecotype	Dune mother	0.465	4.51	0.000
		Headland mother	0.255	1.93	0.074
		All	0.387	4.39	<0.0001
	Ecotype*Region	Dune mother	0.075	0.73	0.474
		Headland mother	-0.041	-0.31	0.760
		All	0.014	0.16	0.877
	Divergence Time	Dune mother	0.000	2.89	0.010
		Headland mother	0.000	0.68	0.506
		All	0.000	2.41	0.021

Continuation Table 4.3. Effects of region, ecotype, their interaction, and of divergence times on the RI for three reproductive barriers and the cumulative between population comparisons in *Senecio lautus*.

Barrier	Effects	Comparison	Estimate	T ratio	*P*
F1 inviability	Region	Dune mother	0.284	1.63	0.130
		Headland mother	0.116	0.71	0.490
		All	0.187	1.52	0.139
	Ecotype	Dune mother	0.172	2.1	0.057
		Headland mother	0.018	0.23	0.822
		All	0.091	1.55	0.131
	Ecotype*Region	Dune mother	0.207	2.52	0.027
		Headland mother	0.132	1.65	0.126
		All	0.185	3.16	0.004
	Divergence Time	Dune mother	0.000	-1.2	0.252
		Headland mother	0.000	-0.08	0.940
		All	0.000	-0.83	0.412
Cumulative	Region	Dune mother	-0.334	-0.9	0.378
		Headland mother	0.207	0.64	0.526
		All	-0.063	-0.26	0.800
	Ecotype	Dune mother	0.462	2.48	0.021
		Headland mother	0.129	0.8	0.433
		All	0.296	2.36	0.022
	Ecotype*Region	Dune mother	0.119	0.64	0.529
		Headland mother	-0.135	-0.84	0.412
		All	-0.008	-0.06	0.950
	Divergence Time	Dune mother	0.000	1.37	0.183
		Headland mother	0.000	0.79	0.438
		All	0.000	1.54	0.131

Table 4.4. Cumulative RI for the different reproductive barriers between coastal ecotypes in *Senecio lautus.*

Region	Ecotype	Mean RI	S.E
Between	Different	0.941	0.261
Between	Same	0.015	0.306
Within	Different	-0.199	0.314
Within	Same	-0.493	0.394

Figures

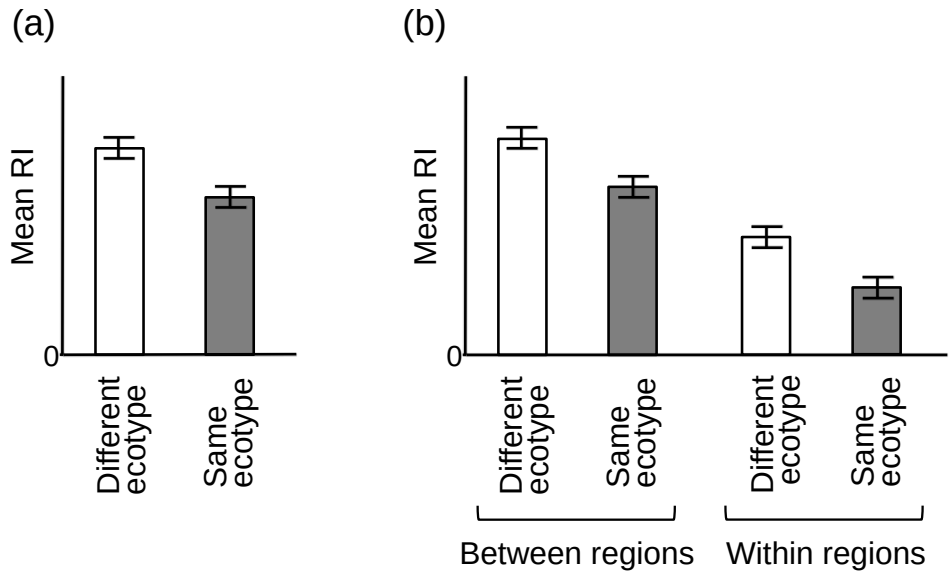

Fig. 4.1. Predictions for the strength of RI in a) a classical model of ecological parallel evolution, and b) in a model that includes two geographical regions (clades) within the system.

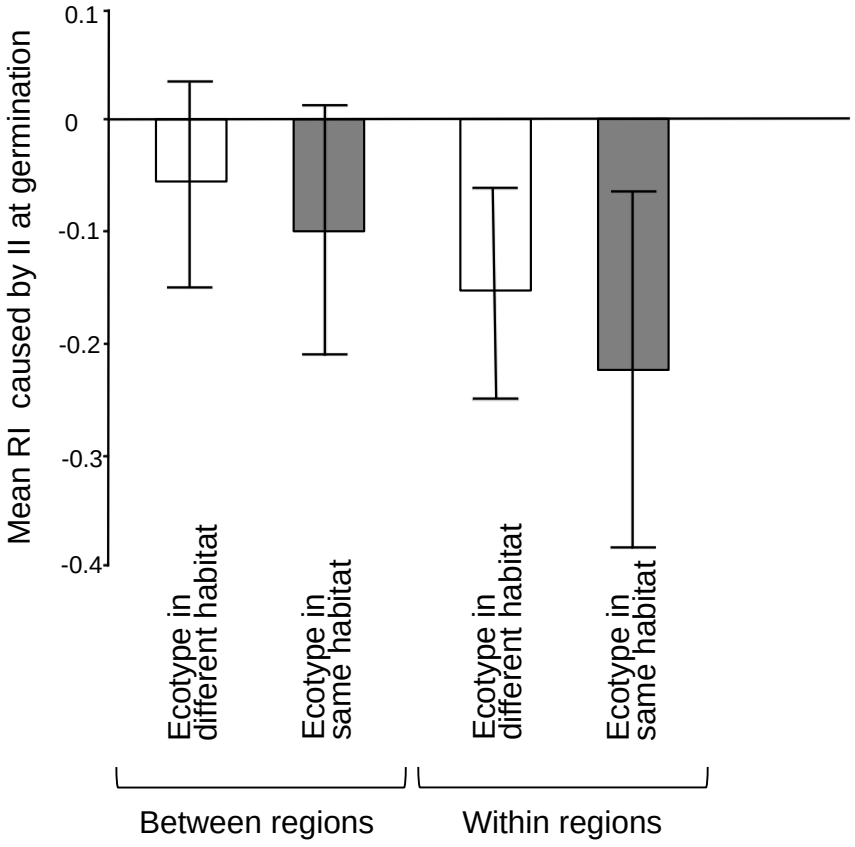

Fig. 4.2. The strength of reproductive isolation (RI) due to immigrant inviability (II) at germination for coastal populations in *Senecio lautus*. RI was estimated when seeds were sowed in soil from the respective type of habitat (Ecotype in same habitat), and when they were sowed at the alternative habitat (Ecotype at different habitat), both within region and between regions.

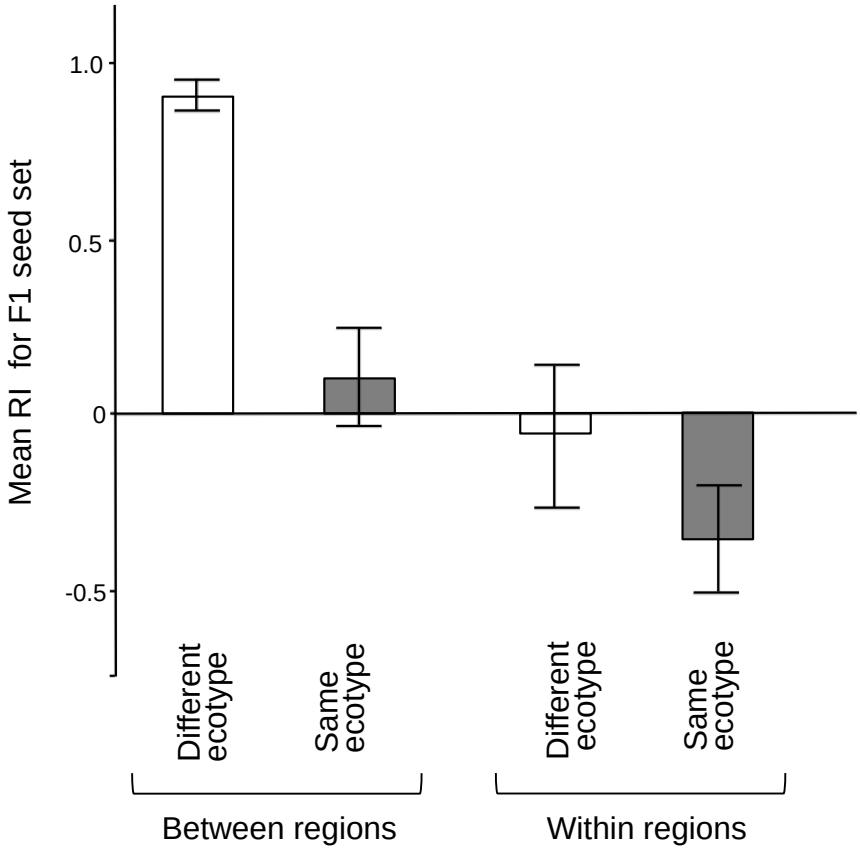

Fig. 4.3. The strength of reproductive isolation (RI) due to F1 hybrid seed set in crosses between coastal populations in *Senecio lautus*. RI was estimated for crosses between populations of the same ecotype and of different ecotypes, both within region and between regions.

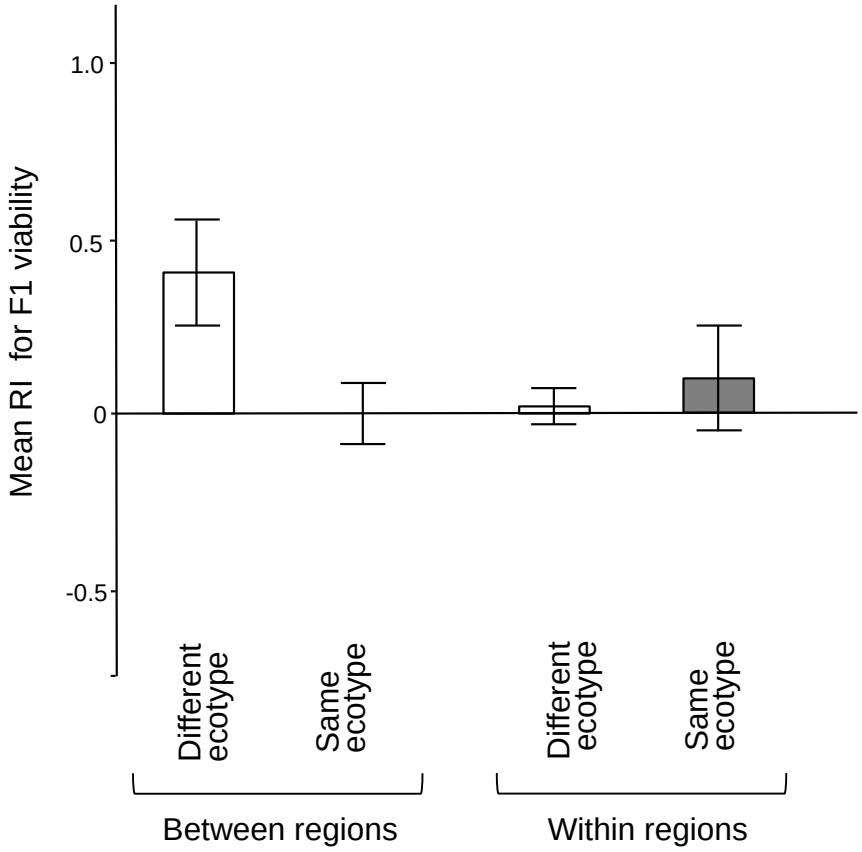

Fig. 4.4. The strength of reproductive isolation (RI) due to hybrid inviability in crosses between coastal populations in *Senecio lautus*. RI was estimated for hybrids produced in crosses between populations of the same ecotype and of different ecotypes, both within region and between regions.

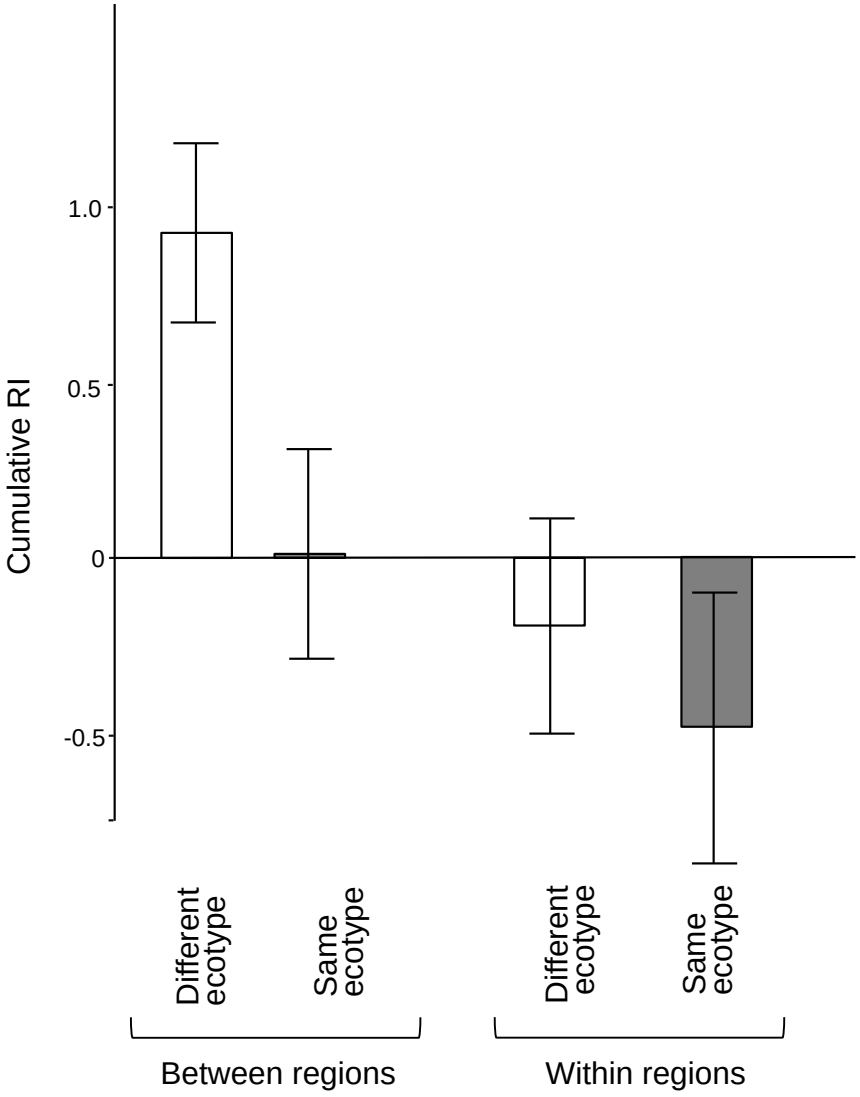

Fig. 4.5. The strength of cumulative reproductive isolation (RI) of the three reproductive barriers in the coastal system in *Senecio lautus.* The cumulative RI was estimated from the sum of the relative RI of immigrant inviabiliy, F1 seed set, and hybrid inviability.

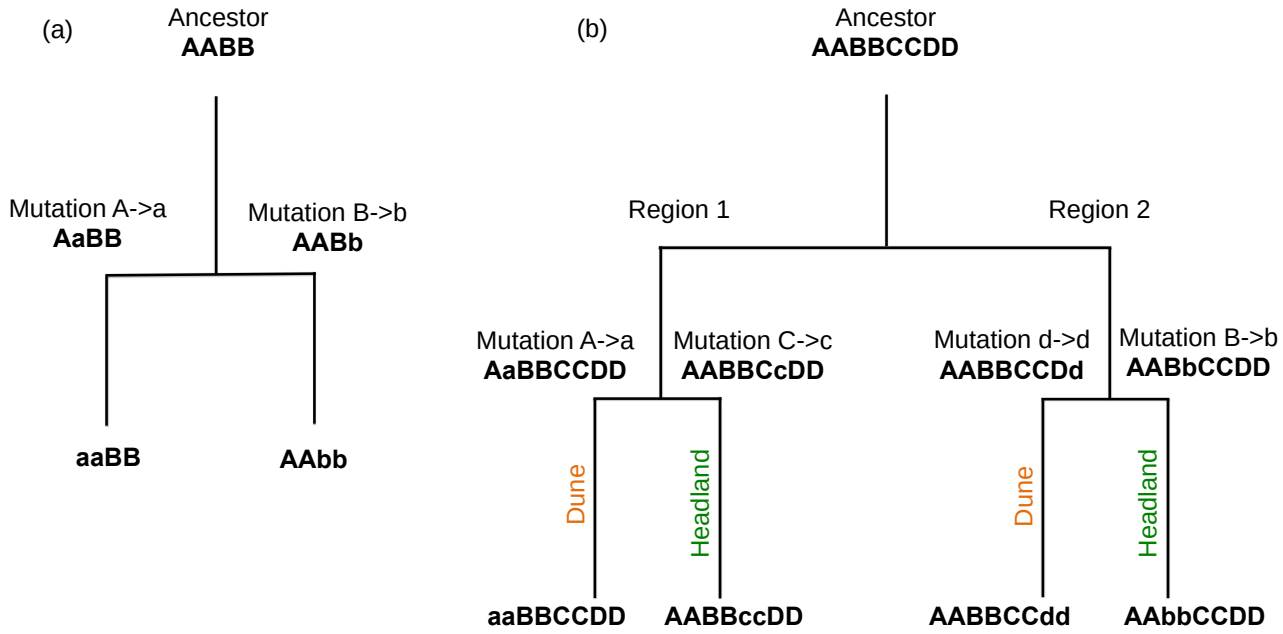

Fig. 4.6. Models of genetic incompatibility. a) Classic Dobzhansky-Muller model of genetic incompatibilities. Hybrids exhibit genetic incompatibilities due to incompatibilities between alleles **a** and **b**: aaBB * AAbb = AaBb. b) Modified model of genetic incompatibilities including four alleles, two ecotypes (Dune and Headland) and two geographic regions. Under this model there are three possible hybridization event from which following genotypes could result: 1) Different ecotype within region: aaBBCCDD*AABBccDD = AaBBCcDD → compatible, and AABBCCdd*AAbbCCDD = AABbCCDd → compatible. 2) Different ecotype between region: aaBBCCDD*AAbbCCDD = AaBbCCDD → incompatible, and AABBccDD*AABBCCdd = AABBCcDc →incompatible. 3) Same ecotype between region: aaBBCCDD*AABBCCdd = AaBBCCDd → compatible, and AABBccDD*AAbbCCDD = AABbCcDD → compatible.

CHAPTER V

GENERAL DISCUSSION AND CONCLUSIONS

Natural selection is a deterministic process and thus it can lead to repeated patterns of evolution (Coyne & Orr, 2004). In animals, it is common to find the replicated evolution of traits that reproductively isolate related descendants from their ancestral populations, but not from each other, in response to adaptation to ecological conditions (Schluter & Nagel, 1995; Nosil, 2012), a process commonly known as parallel ecological speciation. Contrastingly, evidence for parallel ecological speciation is intriguingly rare in plants (Ostevik *et al.*, 2012), perhaps suggesting that the origin of new species differs between plants and animals. I have shown that natural selection has repeatedly and independently driven the evolution of reproductive isolation (RI) between coastal forms of the groundsel *Senecio lautus*, an herbaceous plant found in Australia and the South Pacific Islands. In my dissertation I found that crosses between populations adapted to different habitats were less compatible than crosses between populations adapted to similar habitats (Chapter IV). The magnitude of intrinsic RI depended on whether the mother of the cross was derived from a headland or a sand dune habitat, but it was independent of phylogenetic origin (Chapter III, IV). I confirmed previous reports that coastal population pairs inhabiting sand dunes and rocky headlands have evolved repeatedly and independently multiple times (Chapter III), and showed that natural selection has created extrinsic RI in the field through local adaptation (Chapter II), and implicating the role of plant herbivory in ecotype formation. Unexpectedly, molecular estimates of RI showed that coastal populations are no longer exchanging genes (Chapter III), suggesting that intrinsic and extrinsic RI could evolve at different rates (Mallet, 2006). My results suggest parallel ecological speciation in plants might not be as rare as previously thought and that speciation driven by natural selection may follow similar routes between plants and animals. Below I discuss these main results and how they have helped our understanding of speciation.

Multiple origins of Dune and Headland ecotypes

The parallel evolution of forms suggests that natural selection is responsible for their similarities, and it is unlikely that the vagaries of genetic drift would lead to repeated patterns of evolution. Dune and Headland ecotypes are a system where growth habit (erect Dune and prostrate Headland)

has repeatedly evolved in multiple populations occupying the coast of Australia. Although previous results from Roda *et al.* (2013) have suggested this, I was able to provide an independent assessment of this observation, bringing further credibility to the claim. But, can we really know how many times have Dune and Headland populations evolved independently? And can we know whether they did it *in* situ or after secondary contact? Perhaps, the easiest answers is that they have evolved at least twice, one time in the eastern coasts, and one time in the southern coasts of Australia. I say this because there is deep phylogenetic structure between populations form the two regions (Chapter III), STRUCTURE analyses also detects such division (Chapter III), and previous results from genome-wide scans of variability and from microsatellite data (Roda *et al.* 2013 a,b), all suggest that there are two major clades in the system. These two clades are separated by more than 500 km in distance, and are consistent with the idea that genetic neighbourhoods in *Senecio lautus* are of such length. For instance, the relationship between geographic distance and divergence time (as estimated by IM models) plateaus around such distance between populations, suggesting that divergence saturates beyond this point. It is likely that such saturation results from the initial divergence of the eastern and southern clades and because they have remained isolated from each other over the past (~250 to 500 thousand years ago). Thus, I would suggest that we can think of *Senecio lautus* as a species complex with two major divergence systems geographically isolated from each other, but with migration-drift balance within each region (note that there is strong IBD in each clade and across clades, Chapter III and Roda *et al.,* 2013a).

Isolation by distance in population system suggests that there is some balance between the differentiating forces of genetic drift and the homogenising effects of gene flow. Theories of parapatric speciation suggest that IBD, as well as patchy distribution of habitats, would favour the evolution of species *in situ*. In other words, in such a scenario, gene flow mostly affects local divergence, and local adaptation acts independently amongst nascent population pairs. However, more information would be needed to truly claim that there are multiple instances of Dune and Headland evolution within each region. Some previous data suggests that this may be the case. Roda *et al.* a2013) found that linked sites to outlier loci grouped by geography and not by ecotype. In other words the genetic information contained in the background where favoured mutations occur differs between populations and not between ecotypes. If they differed between ecotypes (i.e., the genetic background of Dunes were the same, and the background of Headlands were the same), then linked sites would cluster populations according to their habitat and not according to their locality. Different genetic backgrounds can exist because of two main reasons. First, genetic background can differ between population pairs because they are evolving independently form each other and gene flow is not strong enough to homogenising them across geography. Alternatively,

they are just standing genetic variation that either differs between populations (due to incomplete lineage sorting) or the similar between populations, but selective mutations are recruited differently in each habitat. Regardless of the scenario, each one of them implicates independence between population pairs, be that from geographic isolation or from recruitment of different alleles form standing genetic variation across geography. In my dissertation I take the cautious approach and I work under the assumption that there are at least two independent origins for Dune and Headland populations, thus creating the rationale for the parallel ecological speciation test shown in Chapter IV. Parallel evolution of forms does not imply parallel speciation between forms. To connect the evolution of traits with the evolution of RI other lines of evidence are required. In turn, I will discuss these criteria, and will evaluate whether the body of work presented here supports the thesis that *Senecio lautus* is a case of parallel speciation by natural selection.

The evolution of intrinsic RI in Senecio lautus

Parallel ecological speciation predicts that populations adapting to similar environments will remain reproductively compatible compared to populations adapting to contrasting environments. To perform this test one requires independent instances of ecotype evolution, as it happens in the coastal population of *S. lautus*. I found two major results for the evolution of RI in the system: Dune and Headland pairs produced fertile hybrids when I crossed individuals from the same region and regardless of ecotype. In contrast, crosses between individuals from different regions were compatible if they were derived from the same ecotype, but were reproductively incompatible if they were derived from different ecotypes. Note that in the first case, gene flow has likely happened in parapatry (Roda *et al.,* 2013a), but has not occurred between regions. It is therefore possible that lack of intrinsic RI within regions has been constrained by gene flow. Alternatively, population divergence between Dune and Headland populations might be too young (Dune – Headland pairs are younger than 200,000 y, chapter III) and there has not been time to accumulate Dobzhansky-Muller incompatibilities in the system. These two explanations are not mutually exclusive, as gene flow will always antagonise the evolution of RI.

Why do young populations adapting to contrasting environments not evolve intrinsic RI? Results from my reciprocal transplant experiments may also provide answers to this question. As in many experiments in the field, F1 hybrids between Dune and Headland parents showed hybrid vigour. If F1 individuals remain alive in the alternative population, the potential for backcross production and therefore introgression increases. Thus, genetically, the way parental alleles interact in a hybrid might be important for the potential for gene flow between populations. Although I do not have data

for testing this idea, it is likely that dominance and epistatic relations manifest differently in hybrids compared to parents, something clearly observed in other systems like *Drosophila* species. However, it is important to note that in my experiment hybrid vigour interacted with development as local types survived better than hybrids at later stages of development. Thus, the effects of hybrid vigour might be more complex than previously anticipated and might require further study in other systems.

Altogether, lack of intrinsic RI and possible hybrid vigour support that speciation, if happening at all, is at the very early stages within regional coasts (note that I have not tested for reductions in hybrid viability and fertility at later hybrid generations such as F2 hybrids where recessive-by-recessive genetic incompatibilities are more likely to manifest.) It is possible that genetic incompatibilities affecting F1 hybrids have not accumulated between parapatric pairs and that gene flow constrains their differentiation beyond morphology. In other words, Dune and Headland populations are clearly ecotypes separated by strong extrinsic RI in the field. However, as mentioned before, although intrinsic RI is overall absent amongst diverging pairs and between populations of the same region, it is present in crosses between populations from different geographic regions. This suggests that genetic incompatibilities might require longer periods of time to accumulate and that later stages of speciation are occurring in the system, but in the absence of gene flow between the major clades (but see section below on patterns of gene flow within clades).

Natural selection creates prezygotic and postzygotic reproductive isolating barriers

Reciprocal transplant experiments in the field suggested that ecological factors in the sandy dunes and rocky headlands produce strong extrinsic RI. I found that the same mechanisms creating RI before mating (immigrant inviability), also creates barriers after hybrids are formed. This is consistent with predictions of a positive relation between immigrant inviability and extrinsic postzygotic isolation (Rundle & Whitlock, 2001; Nosil *et al.*, 2005), but contrasts with empirical evidence in plants that found that extrinsic barriers after mating could be rather weak (Lowry *et al.*, 2008; Schemske, 2010). There are two reasons to think that extrinsic postzygotic isolation could be more common than previously found: First, extrinsic RI has been rarely studied in many plant systems, with reciprocal transplant experiments rarely including hybrids. For instance, Schemske in (2010) found that none of the plant systems in a comparative analysis evaluating the contributions of prezygotic and postzygotic barriers had tested for extrinsic postzygotic isolation. Second, reciprocal transplants that included hybrids usually transplanted individuals at advanced stages of

their life cycle, possibly missing the earliest stages of individual's life cycles (e.g. few days old seedlings). Although I found that immigrant inviability was stronger than extrinsic postzygotic isolation in *S. lautus* (at least in Cabarita beach pair), this could be a consequence of hybrid vigour at germination mitigating the strength of the postzygotic form. It is also possible that for parapatric plant systems extrinsic postzygotic isolation is inappropriately classified as a late acting barrier. In systems like *S. lautus,* where hybrids can be directly created by cross-pollination between habitats, seams reasonable that selection against migrants and hybrids occur at the same time. Therefore, the sequential order in which the relative contribution of reproductive isolating barriers is estimated could underestimate the impact that this extrinsic barrier could have reducing gene flow. The mechanisms driving the evolution of extrinsic RI are challenging to discover, but my results suggests that interactions between plants and herbivores could be important during the early stages of differentiation. Note that it is possible that extrinsic reproductive barriers may no longer be important for the current circumstances where gene flow is no longer present between paratric population pairs.

Predation as a source of divergent natural selection

I found that biotic interactions between herbivores and plants could create strong RI in the coastal system of *S. lautus*. These results differ from other plant systems that have mostly found abiotic interactions (e.g. soil traits; Lowry *et al.*, 2009; Kay *et al.*, 2011) creating divergent natural selection, but resemble cases in animal systems where predator-pray interactions have been found playing this role. It is possible that similar or different herbivores also create divergent natural selection in other parapatric pairs (ants cutting and carrying away entire seedlings have been observed at experimental plots in recent field transplant experiments, Ortiz-Barrientos personal communication). This finding suggests that coastal populations might have evolved adaptive compounds that protect them from the local fauna (as found in other *Senecio* species). This mechanism deserves further investigation through enclosure and other experiments that permit the study of predation isolated from the effect of other habitat traits, and the identification of the adaptive compound creating RI.

Other causes of extrinsic RI

Soil is one of the most contrasting features between the sandy dune and rocky headland environments. However its role creating divergent natural selection between Dune and Headland

populations at the germination stage seams weak. In all of the experiments in which I isolated the effect of soil conducting transplant experiments under controlled conditions, I found that it did not have an effect on the germination of the local parapatric pairs (e.g. Dune and Headland seeds form Cabarita beach germinated equally in the dune and headland soil form Cabarita beach). Although a similar result was obtained in the transplants at the field, a not significant pattern in the direction of local adaptation in this experiment could result from the inclusion of other environmental conditions present in the natural conditions that may interact with soil (e.g. temperature, soil depth, pathogens in the soil, etc.). It is possible that soil could have a weak effect on the total RI of the system that I have did not detect with our current experimental population sizes. It also deserves more attention the effect that other locality soils could have on the germination of coastal pairs. For example, when I tested for Cabarita beach seeds on Stradbroke Island soils, I found significant differences in the germination of ecotypes, and in the direction of local adaptation (Chapter II; Melo *et al.* 2014). However, in chapter IV I found mixed results, with some populations showing patterns of local adaptation (as in Chapter II), but also showing even higher germination at other locality soils. This last result could be owed to a strategy of maximizing colonization, as has been suggested for some weedy plants. It is also possible that soil has an effect at later stages of individual's life cycles, more studies are needed to address this point.

Lack of gene flow between parapatric populations

Estimates of gene flow between Dune and Headland populations suggest that strong extrinsic RI could be causing drastic reductions of it. In Chapter II, I found that from the whole set of reproductive barriers measured in Cabarita beach populations; flowering time, immigrant inviability and extrinsic hybrid viability were the main contributors to the total RI in the system (0.88 in the sandy dunes and 0.76 in the rocky headland). Our direct estimates of gene flow indicate that Dune and Headland do not show recent signatures of gene flow. This suggests that either extrinsic RI is stronger than what could be detected in our experiments, or that an extra barrier could be completing the RI between Dune and Headlands. It is possible that conspecific pollen precedence (CPP), common in plant systems could contribute to the isolation of Dune and Headland. However, it is also possible that extrinsic RI is strong enough to cease gene flow between parapatric pairs, as has been found in other systems (Baldwin, 2005). An alternative explanation is that Dune and Headland populations are effectively allopatric populations that have never experienced gene flow (which opposes the signatures of gene flow found in Roda et al., 2013). In other words, that although these populations are found proximate to each other –sometimes separated by less that 5m– seed dispersal and cross-pollination do not occur. However, this scenario seams rather

unlikely: seeds from *Senecio* are very light and present a morphology conducive for wind dispersal, and importantly, the fact that gene flow was detected between populations separated by ~170 km strongly suggest dispersal can occur over long distances. Another alternative includes an initial divergence in allopatry, followed by gene flow after secondary contact (explaining genomic signatures of gene flow; Roda *et al.,* 2013), although this scenario opposes our result of lack of recent gene flow. Until CPP is tested for our system, we propose that natural selection has driven the divergence of each Dune and Headland pair, and that ecology based RI is strong enough to cease gene flow. In Chapter III we tested for the possible scenarios I have describe, applying novel test that uses patterns of accumulation of polymorphisms. Results from this test where consistent with a scenario of ecological speciation with gene flow, with an effect of strong natural selection isolating whole populations genomes. Therefore we ask again why is that intrinsic RI has not evolved yet between populations if the conditions for genetic drift to act are given. This may not be a novel question, other studies have found that genetic incompatibilities may take a long time to evolve (Mallet, 2006; Levin, 2012), reason why we find this unsurprising, spatially in young systems as *S. lautus.*

Senecio lautus in the progress toward speciation

If natural selection can overcome gene flow at early stages of speciation is not under the spotlight of speciation theory any more. Good examples have demonstrated that divergent natural selection can promote divergence, taking populations to different degrees of differentiation in the speciation continuum. For example in sticklebacks, stream - lake transitions show varying degrees of populations divergence, from weak genetic and morphological differentiation, to populations at advanced stages differentiation. The controversy is now focused on whether divergent natural selection is able to complete speciation. In other words, weather natural selection is only required to restrict gene flow to the point in which genetic drift could favour the evolution intrinsic RI (Thibert Plante & Hendry, 2010). Intrinsic RI has been thought to assure the persistence of species even under an eventual case of habitat disturbance. However, theory of allopatric speciation suggests that intrinsic reproductive barriers may not prevent populations from fusion, as may occur if prezygotic isolation does not evolve after secondary contact, in which case intrinsic RI could be reversed. This leads us to think if intrinsic RI is necessary to consider speciation complete. Researches on a wide variety of organisms have suggested that intrinsic postzygotic isolation evolves at a slower pace than species do. For example it has been suggested that birds speciate at the same pace as mammals, although they evolve hybrid inviability much slower (~21 million y, Avise *et al.*, 1998). Similarly, young butterfly species have show little of no intrinsic RI (Mallet, 2006), and in fish species strong

sexual selection evolves before any intrinsic RI (Mendelson, 2003). In plants, studies have found that hybrid sterility evolves with increasing genetic distance (Moyle *et al.*, 2004), Levin (2012) found that considerable reductions in hybrid fertility tend to evolve after more than 4 million years. This independently to weather intrinsic RI evolves in allopatry or sympatry (Mallet, 2006). Despite this, there seams to be an agreement in the field to consider that speciation requires intrinsic RI, producing species that would persist in time despite eventual habitat disturbance. For example, the test of parallel ecological speciation includes forms of intrinsic RI to consider that evidence for two of its criteria is strong. Also various definitions of the stages of the speciation continuum (Clausen, 1951; Wu, 2001; Hendry, 2009) require the evolution of 'irreversible' RI at the last stage of the process. This commonality is likely due to the fact that intrinsic RIs have been considered irreversible barriers, and their appearance may assure species to perdure.

Dune and Headland ecotypes in *S. lautus* resemble other cases where fertile populations in the glasshouse, but there is no evidence for hybridization or gene flow but are still considered distinct species (Baldwin, 2005). *Layia discoidea* was even considered from a different genus to *L. glandulosa,* but phylogenetic analysis and crossing experiments found they were sister species. Under the speciation continuum point of view, we learn of speciation as a process in opposition to the final product. Under this view ecotypes are stages of the process, and weather they remain as such, reverse or evolve towards separate species is possible. In the case of Dune and Headland populations it is uncertain if they will evolve to discrete species. With the evidence in the system all we know is that it is an excellent system to study ecological speciation.

Future directions

Senecio lautus has rapidly turned into a promising model to study speciation. Not only presenting evidence for the parallel ecological parallel speciation, but setting the arena to conduct new studies that will allow us to understand speciation in more detail. Throughout the various projects in my dissertation I identified specific aspects that arose directly from my discoveries and that would help us understand ecological speciation in the Dune and Headland system. For example, although we have been able to detect high genetic compatibility between Dune and Headland populations, we can only assert this under controlled conditions and for early hybrid generations (F1). Transplant experiments in the two environments that include hybrids with different contributions of Dune and Headland genomes, and the study of advanced generations can solve this gap. From our experiments in the field, I found that predation could be an important mechanism favouring speciation in plants. This mechanism required further investigation in experiments that isolate the

effect of predation from other environmental factors and that identifies and test for a possible adaptive trait (compound). Finally, although we have been able to explore the contributions of various RI barriers, finding that extrinsic RI causes strong RI between Dune and Headland populations, lack of gene flow could suggest there could be other barriers acting in the system. It is possible that conspecific pollen precedence a postmating prezygotic barriers common in several plant species, could be contributing to the complete isolation of Dune and Headland populations. Finally, our results suggest that natural selection could have a larger effect isolating the genomes of populations than previously expected. More work is needed on aspects such as inversions, recombination rates and mutation rates, which would bring insight onto which are the genetic architectures that favour the process of ecological speciation with gene flow. Moreover, identifying the genes responsible for both adaptation and speciation will help us to test the fundamental hypothesis that natural selection causes directly the evolution of reproductive isolation.

References

Avise JC, Walker D, Johns GC. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**: 1707-1712.

Baldwin BG. 2005. Origin of the serpentine-endemic herb *Layla discoidea* from the widespread *L. glandulosa* (Compositae). *Evolution* **59**: 2473-2479.

Coyne JA, Orr HA. 2004. *Speciation*: Sinauer Associates Sunderland, MA.

Clausen J. 1951. Stages in the evolution of plant species. *Stages in the evolution of plant species.*(6d).

Hendry APHAP. 2009. Ecological speciation! Or the lack thereof? . *Canadian Journal of Fisheries and Aquatic Sciences* **66**: 1383-1398.

Kay KM, Ward KL, Watt LR, Schemske DW. 2011. Plant speciation. *Serpentine: the evolution and ecology of a model system. University of California Press, Berkeley*: 71-96.

Levin DA. 2012. The long wait for hybrid sterility in flowering plants. *New Phytologist* **196**: 666-670.

Lowry D, Hall M, Salt D, Willis J. 2009. Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland Mimulus guttatus. *New Phytologist* **183**: 776-788.

Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 3009-3021.

Mallet J. 2006. What does Drosophila genetics tell us about speciation? *Trends in Ecology & Evolution* **21**: 386-393.

Mendelson TC. 2003. Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: Etheostoma). *Evolution* **57**: 317-327.

Moyle LC, Olson MS, Tiffin P. 2004. Patterns of reproductive isolation in three angiosperm genera. *Evolution* **58**: 1195-1208.

Nosil P. 2012. *Ecological speciation*: Oxford University Press.

Nosil P, Vines T, Funk D. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**: 705-719.

Ostevik KL, Moyers BT, Owens GL, Rieseberg LH. 2012. Parallel ecological speciation in plants? *International Journal of Ecology* **2012**.

Roda F, Ambrose L, Walter GM, Liu H, Schaul A, Lowe A, Pelser PB, Prentis P, Rieseberg LH, Ortiz-Barrientos D. 2013a. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Molecular Ecology* **22**: 2941-2952.

Roda F, Liu H, Wilkinson M, Walter G, James M, Bernal D, Melo M, Lowe A, Rieseberg L, Prentis P, Ortiz-Barrientos D. 2013b. Convergence and divergence during the adaptation to similar invironments by an Australian groundsel. *Evolution* **67**: 2515-2529.

Rundle HD, Whitlock MC. 2001. A genetic interpretation of ecologically dependent isolation. *Evolution* **55**: 198-201.

Schemske DW. 2010. Adaptation and The Origin of Species. *American Naturalist* **176**: S4-S25.

Schluter D, Nagel LM. 1995. Parallel Speciation by Natural Selection. *The American Naturalist* **146**: 292-301.

Thibert Plante X, Hendry A. 2010. When can ecological speciation be detected with neutral loci? *Molecular Ecology* **19**: 2301-2314.

Wu CI. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**: 851-865.

SUPLEMENTARY MATERIAL

SUPPLEMENTARY CHAPTER II

Table S2.1. Geographic location of populations used in this study.

Locality	Environment	Location
Cabarita Beach	Dune	S 28° 19' 54.66" E 153° 34' 17.04"
	Headland	S 28° 21' 45.07" E 153° 34' 46.82"
Hat Head	Dune	S 30° 52' 57.06" E 153° 4' 7.32"
	Headland	S 30° 53' 15.06" E 153° 4' 3.96"
Lennox Head	Dune	S 28° 47' 10.7" E 153° 35'
	Headland	S 28° 48' 10" E 153° 36' 9.94"
Bayron Bay	Headland	S 28° 38' 4.62" E 153° 38' 14.22"
Lamington Park	Tableland	S 28° 13' 49.83" E 153° 08' 6.28"
Port Macquarie	Dune	S 31° 28' 31.14" E 152° 56' 14.46"
Stradbroke Island	Dune	S 27° 23' 20.20" E 153° 27' 13.71”
	Headland	S 27° 26' 9.94" E 153° 32' 42.81”

Table S2.2. Proportion of individuals that germinated in Dune or Headland soil under glasshouse (GH) or field (FIELD) conditions. The analysis was performed pooling together both F1-D and F1-H into a single category F1. F1-D are hybrids with a Dune cytoplasm and F1-H are hybrids with a Headland cytoplasm (CB – Cabarita Beach, LN – Lennox Head, SI – Stradbroke Island).

Experiment	Soil/Habitat	Genotype	N	G	Std Err
GH	Dune	D	160	0.42	0.039
		H	160	0.32	0.037
		Mean	320	0.37	0.027
	Headland	D	160	0.25	0.035
		H	160	0.20	0.032
		Mean	320	0.23	0.024
	Dune(CB)	D	35	0.31	0.080
		H	35	0.23	0.072
		Mean	70	0.27	0.053
	Headland(CB)	D	35	0.49	0.086
		H	35	0.43	0.085
		Mean	70	0.46	0.060
	Dune(LH)	D	35	0.48	0.086
		H	35	0.06	0.040
		Mean	70	0.27	0.053
	Headland(LH)	D	35	0.14	0.060
		H	35	0.48	0.086
		Mean	70	0.31	0.055
	Dune(SI)	D	35	0.46	0.085
		H	35	0.23	0.072
		Mean	70	0.34	0.057
	Headland(SI)	D	35	0.31	0.080
		H	35	0.26	0.07
		Mean	70	0.28	0.054

Continuation Table S2.2. Proportion of individuals that germinated in Dune or Headland soil under glasshouse (GH) or field (FIELD) conditions.

Experiment	Soil/Habitat	Genotype	N Rows	Proportion Germinated	Std Err
FIELD	Dune	D	275	0.34	0.029
		F1-D	159	0.54	0.040
		F1-H	159	0.40	0.039
		H	269	0.30	0.028
		F1 *	318	0.47	0.028
		Mean	862	0.38	0.016
	Headland	D	271	0.12	0.020
		F1-D	151	0.22	0.034
		F1-H	153	0.23	0.034
		H	265	0.18	0.024
		F1 *	304	0.23	0.024
		Mean	840	0.18	0.013

Table S2.3. Mortality in the field as measured by the average number of days alive at the time of flowering, and the proportion of individuals (seedlings) killed by herbivores in the Dune and Headland habitats.

Measurement	Habitat	Genotype	N Rows	Average days alive/ proportion killed	Std Err
SURVIVAL	Dune	D	94	115.78	11.15
	Dune	F1-D	86	65.10	9.82
	Dune	F1-H	64	75.13	12.26
	Dune	H	83	57.61	9.36
	Dune	F1 *	150	69.38	7.67
	Headland	D	34	30.77	8.02
	Headland	F1-D	34	41.18	8.16
	Headland	F1-H	36	47.11	8.29
	Headland	H	52	46.52	7.06
	Headland	F1 *	70	44.23	5.79
PREDATION	Dune	D	94	0.27	0.04
	Dune	F1-D	86	0.53	0.05
	Dune	F1-H	64	0.48	0.06
	Dune	H	82	0.52	0.05
	Headland	D	32	0.53	0.09
	Headland	F1-D	34	0.53	0.09
	Headland	F1-H	35	0.37	0.09
	Headland	H	44	0.29	0.07

* The analysis was performed pooling together both F1-D and F1-H into a single category F1

Table S2.4. List of statistical tests conducted in the experiments of Melo et al.

Test	Locality	Comparison	Statistical test	N	Parameter	P value
F1 seed set	Glasshouse	Parental (intrapopulation) crosses Vs. interpopulation crosses	ANOVA	64	F = 0.0825	0.9693
Seed germination	Glasshouse - D Soil	Dune population seeds Vs. Headland population seeds	GLM	320	z = -1.732	0.0834
Seed germination	Glasshouse - H Soil	Dune population seeds Vs. Headland population seeds	GLM	320	*z* = -1.059	0.2900
Seed germination	Field	All crosstype germination in the Sandy Dunes Vs. Rocky Headland	GLM	1702	*z* = -9.279	0.0000
Seed germination	Field - Sandy Dunes	Dune population seeds Vs. Headland population seeds	GLM	862	*z* = -0.913	0.3610
Seed germination	Field - Rocky Headland	Dune population seeds Vs. Headland population seeds	GLM	840	*z* = -1.673	0.0943
Seed germination	Field - Sandy Dunes	F1 cross seeds Vs. Dune population seeds	GLM	862	z = 3.137	0.0017
Seed germination	Field - Sandy Dunes	F1-D (D cytoplasm) cross seeds Vs. Dune population seeds	GLM	862	*z* = 3.973	0.0001

Test	**Locality**	**Comparison**	**Statistical test**	**N**	***Parameter***	**P value**
Seed germination	Field - Sandy Dunes	F1-H (H cytoplasm) cross seeds Vs. Dune population seeds	GLM	863	z = 1.248	0.2120
Seed germination	Field - Rocky Headland	F1 cross seeds Vs. Headland population seeds	GLM	840	*z = 1.366*	0.1719
Seed germination	Field - Rocky Headland	F1-D (D cytoplasm) cross seeds Vs. Headland population seeds	GLM	840	z = 1.036	0.3001
Seed germination	Field - Rocky Headland	F1-H (H cytoplasm) cross seeds Vs. Headland population seeds	GLM	840	z = 1.251	0.2108
Survival analyses	Field - Sandy Dunes	Dune population Vs. Headland population	GLM	327	z = -5.847	0.0000
Survival analyses	Field - Sandy Dunes	F1-D (D cytoplasm) cross Vs. Dune population	GLM	327	z = -3.035	0.0024
Survival analyses	Field - Sandy Dunes	F1-H (H cytoplasm) cross Vs. Dune population	GLM	327	z = -3.158	0.0015
Survival analyses	Field - Rocky Headland	Dune population Vs. Headland population	GLM	156	z = -1.396	0.1630

Test	Locality	Comparison	Statistical test	N	Parameter	P value
Survival analyses	Field - Rocky Headland	F1-D (D cytoplasm) cross Vs. Headland population	GLM	156	z = 0.459	0.6460
Survival analyses	Field - Rocky Headland	F1-H (H cytoplasm) cross Vs. Headland population	GLM	156	z = 0.276	0.7820
Predation analyses	Field - Sandy Dunes	Dune population Vs. Headland population	GLM	326	z = 3.753	0.0001
Predation analyses	Field - Sandy Dunes	F1-D (D cytoplasm) cross Vs. Dune population	GLM	326	z = 3.763	0.0002
Predation analyses	Field - Sandy Dunes	F1-H (H cytoplasm) cross Vs. Dune population	GLM	326	z = 2.904	0.0037
Predation analyses	Field - Rocky Headland	Dune population Vs. Headland population	GLM	145	z = 2.159	0.0309
Predation analyses	Field - Rocky Headland	F1-D (D cytoplasm) cross Vs. Headland population	GLM	145	z = 2.124	0.0337
Predation analyses	Field - Rocky Headland	F1-H (H cytoplasm) cross Vs. Headland population	GLM	145	z = 0.766	0.4434
Other mortality causes	Field - Sandy Dunes	Dune population Vs. Headland population	GLM	326	z = 0.235	0.8139
Other mortality causes	Field - Sandy Dunes	F1-D (D cytoplasm) cross Vs. Dune population	GLM	326	z = -0.404	0.6863

Test	Locality	Comparison	Statistical test	N	Parameter	P value
Other mortality causes	Field - Sandy Dunes	F1-H (H cytoplasm) cross Vs. Dune population	GLM	326	z = -0.017	0.9863
Other mortality causes	Field - Rocky Headland	Dune population e Vs. Headland population	GLM	145	z = 0.103	0.9181
Other mortality causes	Field - Rocky Headland	F1-D (D cytoplasm) cross Vs. Headland population	GLM	145	z = -0.038	0.9699
Other mortality causes	Field - Rocky Headland	F1-H (H cytoplasm) cross Vs. Headland population	GLM	145	z = -0.186	0.8526
Fecundity	Field - Sandy Dunes	All cross types	GML	84	F =0.7789	0.5092
Fecundity	Field - Rocky Headland	All cross types	GML	47	F =0.3474	0.7912
Germination of CB seeds in other localities soil	Cabarita Beach Soil – D Soil	CB Dune population Vs. CB Headland population	GLM	70	z = -0.804	0.4210
Germination of CB seeds in other localities soil	Cabarita Beach Soil – H Soil	CB Dune population Vs. CB Headland population	GLM	70	z = -0.305	0.7600
Germination of CB seeds in other localities soil	Lennox Head Soil - D Soil	CB Dune population Vs. CB Headland population	GLM	70	z = -2.794	0.0050

Test	Locality	Comparison	Statistical test	N	Parameter	P value
Germination of CB seeds in other localities soil	Lennox Head Soi – H Soil	CB Dune population Vs. CB Headland population	GLM	70	z = 2.942	0.0030
Germination of CB seeds in other localities soil	Stradbroke island Soi – D Soil	CB Dune population Vs. CB Headland population	GLM	70	z = -1.938	0.0526
Germination of CB seeds in other localities soil	Stradbroke Soi – H Soil	CB Dune population Vs. CB Headland population	GLM	70	z = -0.549	0.5830
Germination of CB seeds in other localities soil	Glasshouse	Full interaction model Vs. simpler one	ANOVA	1df	X^2 = 36.807	0.0000
Morphology (Height)	Glasshouse	Dune population Vs. Headland population	ANOVA	1,37	*F*= 33.501	0.0000
Morphology (No. Branches)	Glasshouse	Dune population Vs. Headland population	ANOVA	1,37	*F*= 28.655	0.0000

Fig. S2.1.

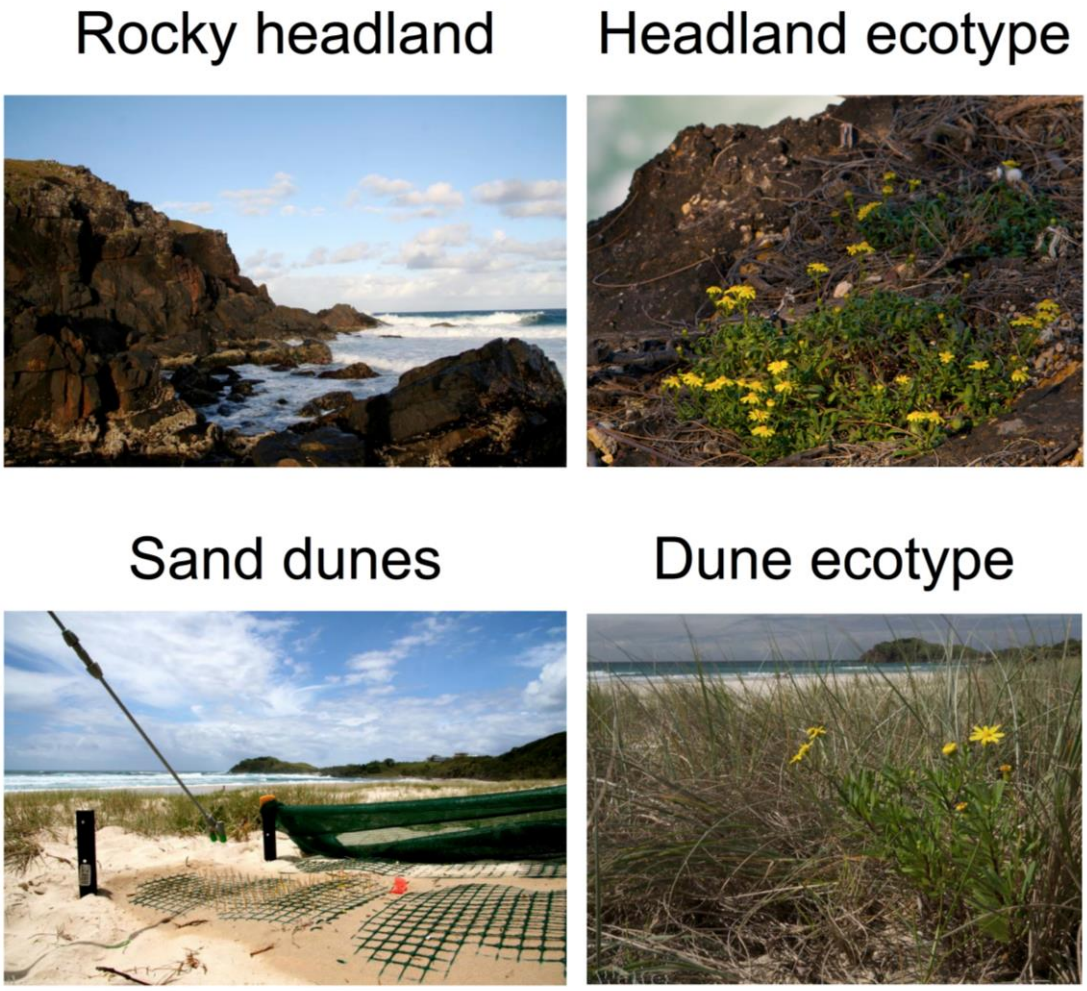

Fig. S2.1. The top pictures show the rocky headland at Cabarita Beach, and a typical Headland plant growing at this site. The pictures below show both the sand dunes at Cabarita Beach and an example of the set up for the transplant experiments, and a Dune individual.

Fig. S2.2.

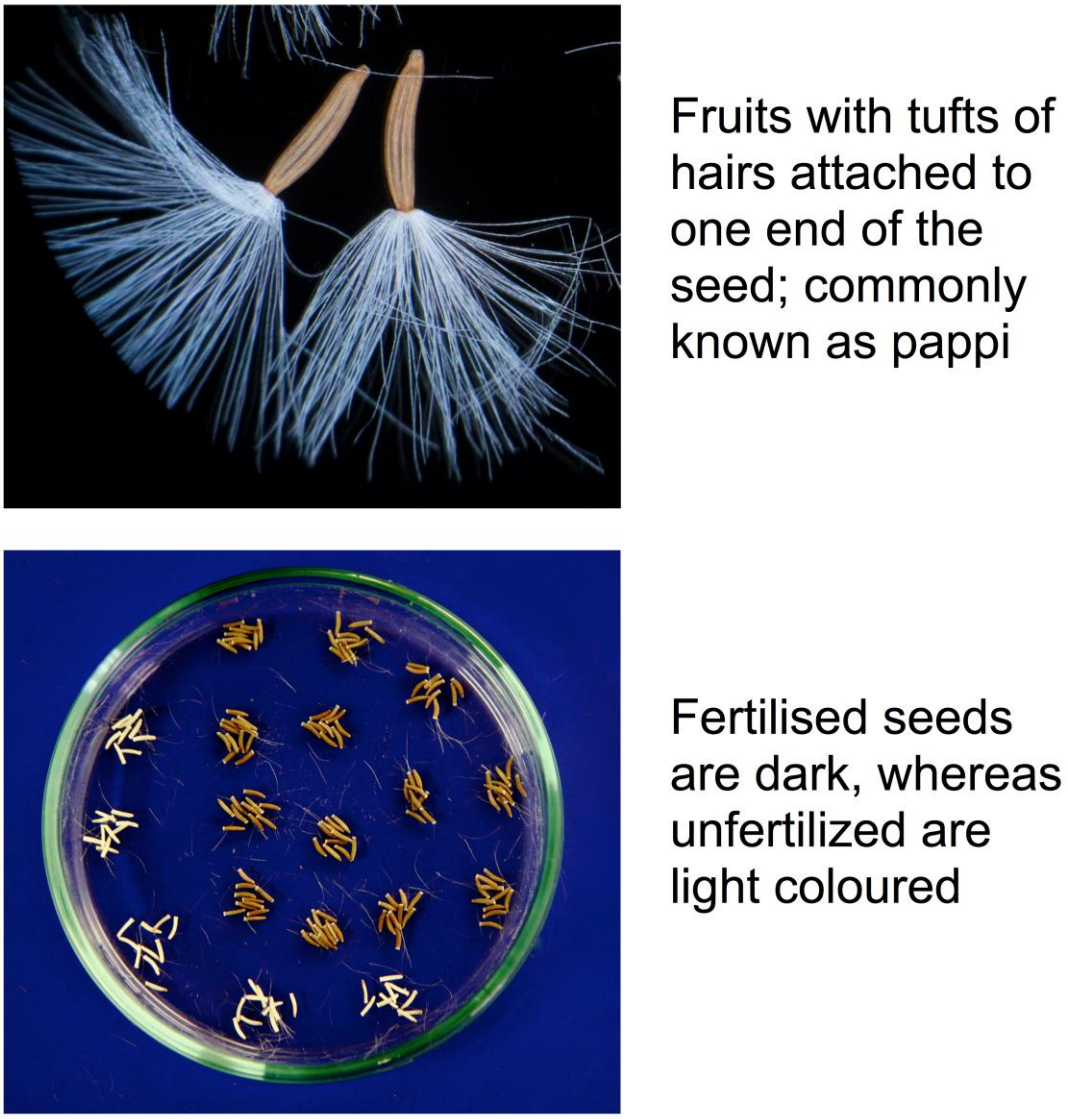

Fig. S2.2. Fruits of *Senecio lautus* (top panel) and an example of seed counting (lower panel) in the Dune population.

Fig. S2.3.

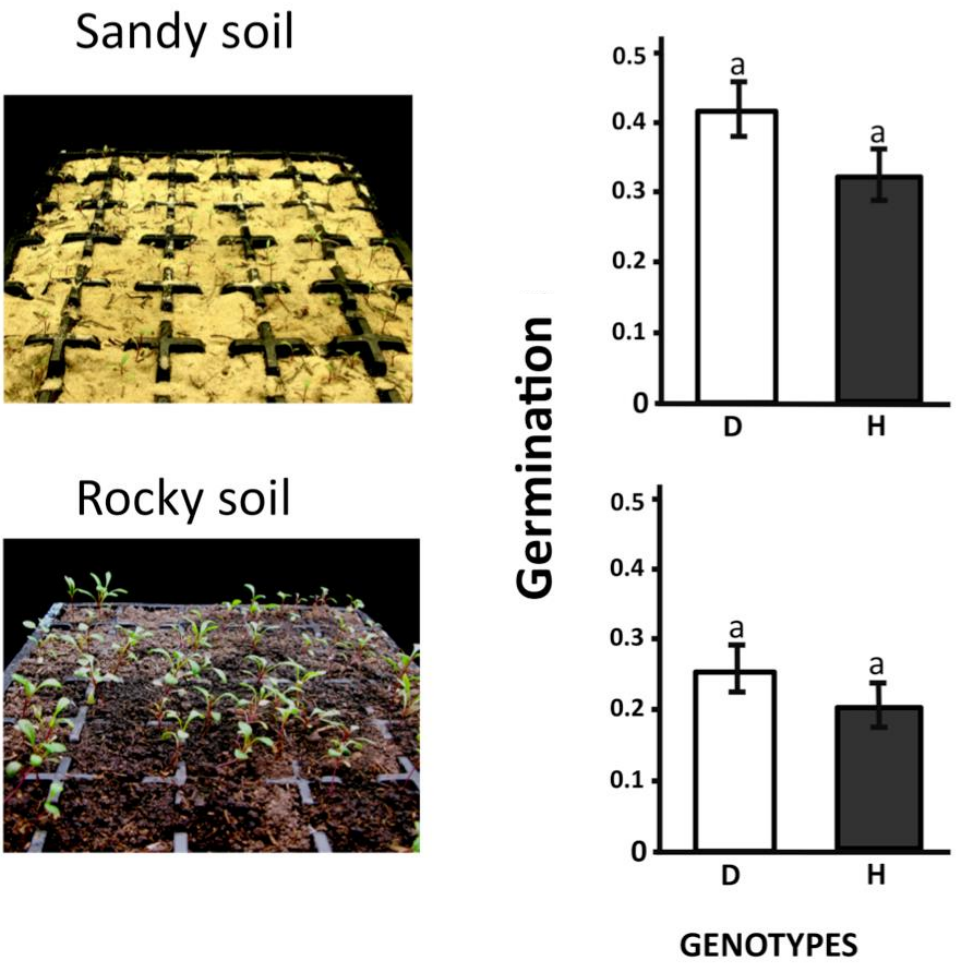

Fig. S2.3. Transplant experiments under controlled conditions. Left pictures show plants growing in trays filled with soil collected from either the sand dunes or the rocky headlands at Cabarita Beach. Right panels show the proportion of Dune and Headland seeds that germinated in each environmental condition. Bars show means and standard errors for binomial probabilities, with different letters denoting significant differences in the proportion of germinations between genotypes by using a nominal logistic model.

Fig. S2.4.

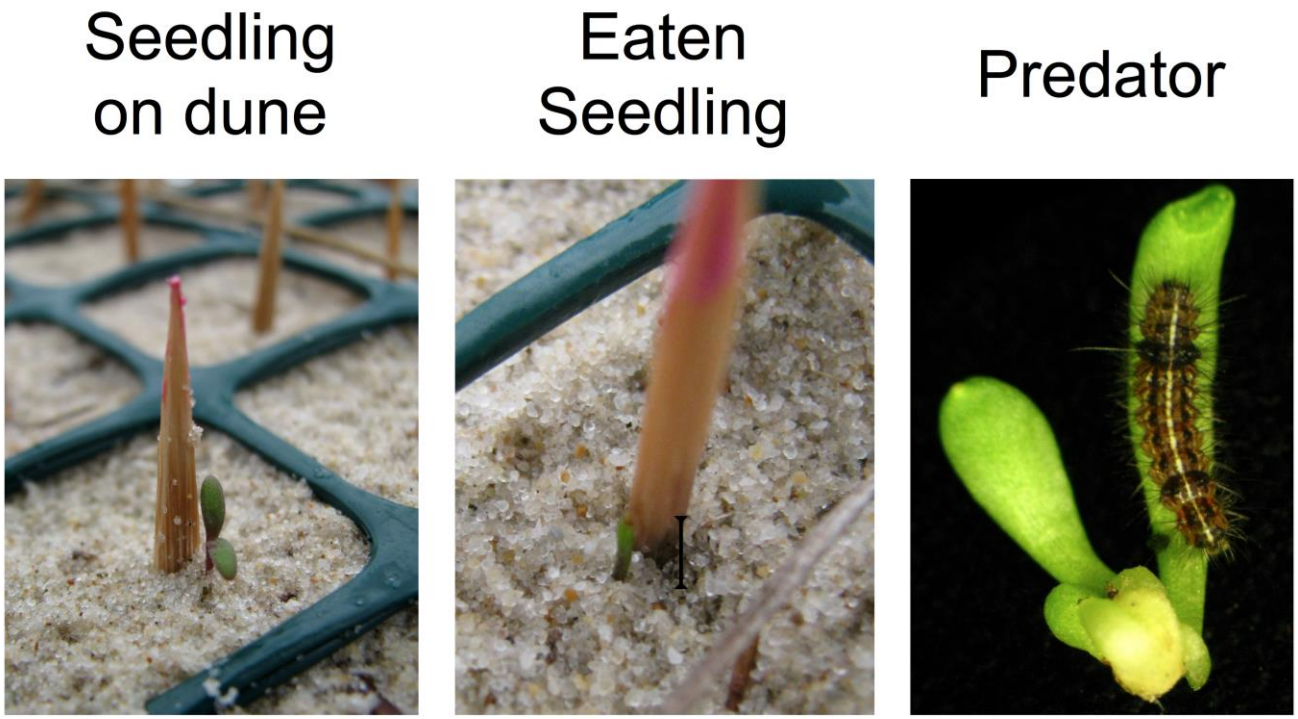

Fig. S2.4. Examples of seedling individual in the field (left panel), a predated seedling (middle panel), and one of the predator individuals (*Spilosoma sp.*) found eating a seedling in the field, and eating a leaf under the stereoscope (right panel).

Fig. S2.5.

Fig. S2.5. Examples of *Senecio lautus* pollinators observed in both sand dune and rocky headlands field.

SUPPLEMENTARY CHAPTER III

Short Access Array amplification: The system uses an Integrated Fluidic Circuit (IFC), which is a microfluidic chip that systematically combines the 48 sample inputs with the 48 primer inputs to create a 2.304 combinations of samples and primers. Three separate machines are involved in the whole amplification process: one pre-PCR to load samples and primers, a thermocycler for amplification, and one post-PCR to harvest the PCR products. Once primers and samples are loaded to the chip, PCR reactions amplify the target-specific regions with common sequence (CS) tags from 50 g of genomic DNA samples in the Acces Array IFC. The harvested PCR products are polled from the Access Array IFC and divide into two following PCR reactions in micrometer plates so the barcodes and sequencing adaptors could be attached. This produced amplicons with a unique barcode by sample, and that also have Ion Torrent PGM sequencing adaptors (A and P1). Tagged amplicons were then ready to be input into the emulsion PCR with Ion Sphere ™ particles and sequencing from both ends of the garget region with a single-read sequencing run.

Table S3.1. 26 loci (L) sequenced for *Senecio lautus*,their location in the genome (Contig) and in the linkage map (C or chromosome number), and neutrality tests (T) for each one of them for each of the populations in the study. A corresponds to the Fu & Li's D*, B to Fu and Li's F*, C to HKY and D to Tajima's D. For each test we report the significance level (with a number 1 in the column) at which each gene signifincantly deviated from neutrallity: ** p<0.01, * p<0.05 and ns p>0.05. NP stand for non-polymorphic loci, M for loci with missing sequences for the outgroup, and error when the program was not able to conduct the test. Genes where HKA test was performed where judged according to its results. For those genes with a missing outgroup (M), we based our desition Fu and Li's tests and Tajima's D results. Genes in light green are neutrally evolving genes. Genes in dark green are neutrally evolving but presenting recombination events. Genes in white deviated from neutral expectations.

L	C	Contig	Test	A03			A05			D01			D14		D23		D03		D32			D04			H01			H12			H15		H02		H21			H05			I01			I02		
				**	*	ns	**	*	ns	**	*	ns	*	ns	*	ns	*	ns	**	*	ns	**	*	ns	**	*	ns	**	*	ns	*	ns	*	ns	**	*	ns	**	*	ns	**	*	ns	**	*	ns
1004		NODE_201534_length_1032_cov_12.200582	A			1			1			1		1		1		1		1		NP	NP	NP			1			1			1		1			1			1			1		
			B			1			1			1		1		1		1		1		NP	NP	NP			1			1			1		1			1			1			1		
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1			1			1		1		1		1		1		NP	NP	NP			1			1			1		1			1			1			1		
1011		NODE_2617158_length_2339_cov_8.104318	A			1	NP	NP	NP			1	NP	NP		1	NP	NP			1			1			1			1			1		1			1			1			1		
			B			1			1			1	NP	NP		1	NP	NP			1			1			1			1			1		1			1			1			1		
			C	NP	NP	NP	NP	NP	NP			1	NP	NP		1	NP	NP			1			1			1			1			1		1	NP	NP	NP			1			1		
			D			1			1			1	NP	NP		1	NP	NP			1			1			1			1			1		1			1			1			1		
1014	1	NODE_1489600_length_1928_cov_8.266598	A			1			1			1	1		1			1			1			1			1		1			1		1			1			1			1		
			B			1			1			1		1	1			1			1			1			1		1			1		1			1			1			1		
			C			1			1			1		1		1		1			1			1			1			1			1		1			1			1			1		
			D			1			1			1		1		1		1			1			1			1		1			1		1			1			1			1		
1018		NODE_1395294_length_3443_cov_9.196050	A			1	NP	NP	NP	NP	NP	NP		1		1		1			1			1			1			1			1		1			1			1			1		
			B			1	NP	NP	NP	NP	NP	NP		1		1		1			1			1			1			1			1		1			1			1			1		
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1	NP	NP	NP	NP	NP	NP		1		1		1	1					1			1			1			1		1			1			1		1		
1021		NODE_1104840_length_662_cov_7.762840	A	1				1		1			1		NP	NP		1			1		1			1			1			1		1			1			1			1		
			B	1				1		1			1		NP	NP		1			1		1			1			1			1		1			1			1			1		
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D		1				1	1			1		NP	NP		1			1		1			1			1			1		1			1			1			1		
1024		NODE_4941699_length_3074_cov_8.074821	A			1			1			1	NP	NP		1		1	1					1			1			1			1		1			1			1			1		
			B		1				1			1	NP	NP		1		1	1					1			1			1			1		1			1			1			1		
			C			1			1			1	NP	NP		1	NP	NP			1			1			1			1			1		1	NP	NP	NP			1			1		
			D			1			1			1	NP	NP		1		1	1					1			1			1			1		1			1			1			1		
1080		NODE_3879988_length_2219_cov_7.630464	A			1			1			1		1		1		1			1	1			NP	NP	NP			1			1		1	NP	NP	NP		1				1		
			B			1			1			1		1		1		1			1	1			NP	NP	NP			1			1		1	NP	NP	NP		1				1		
			C			1			1			1		1		1		1			1			1	NP	NP	NP			1			1		1	NP	NP	NP			1				1	
			D			1		1				1		1		1		1			1	1			NP	NP	NP			1			1		1	NP	NP	NP	1					1		
1084		NODE_1029575_length_1766_cov_10.731030	A			1			1	NP	NP	NP		1		1		1			1		1			1			1			1		1			1	NP	NP	NP			1			1	
			B			1			1	NP	NP	NP		1		1		1			1		1			1			1			1		1			1	NP	NP	NP			1			1	
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1			1	NP	NP	NP		1		1		1			1		1			1			1			1		1			1			1			1			1	
1085		NODE_112497_length_728_cov_14.436813	A			1	1			1				1	1		1			1			1			1			1			1		1			1		1			1			1	
			B			1	1			1				1	1		1			1			1			1			1			1		1			1	1					1		
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1	1				1			1	1		1			1			1			1			1			1		1			1		1			1			1	
1090		NODE_1895850_length_1276_cov_8.061129	A			1			1			1		1	1		1			1		1			1			1		1			1		1	NP	NP	NP			1			1	
			B			1			1			1		1	1		1			1		1			1			1		1			1		1	NP	NP	NP			1			1	
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1			1			1		1	1		1			1		1			1			1		1			1		1	NP	NP	NP			1			1	
1091		NODE_1910551_length_1103_cov_8.417951	A			1			1			1		1		1	1			1			1			1			1			1		1			1			1			1		
			B			1			1			1		1		1	1			1			1			1			1			1		1			1			1			1		
			C			1			1			1		1		1	1			1			1			1			1			1		1			1			1			1		
			D			1			1			1		1		1	1			1			1			1			1			1		1			1			1			1		
1096	16	NODE_1099407_length_1032_cov_11.230620	A			1			1			1	NP	NP	NP	NP	1			1			1			1			1	NP	NP	1		1			1			1			1		
			B			1			1			1	NP	NP	NP	NP	1			1			1			1			1	NP	NP	1		1			1			1			1		
			C			1			1			1	NP	NP	NP	NP	1			1			1			1			1	NP	NP	1		1			1			1			1		
			D			1			1			1	NP	NP	NP	NP	1			1			1			1			1	NP	NP	1		1			1			1			1		
1098		NODE_2932452_length_6364_cov_8.568039	A			1			1		1			1		1	1			1			1			1			1			1		1			NA	NA	NA	1			1		
			B			1			1		1			1		1	1			1			1			1			1			1		1		1	NA	NA	NA	1			1		
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1			1		1			1	1		1			1			1			1			1			1		1		1	NA	NA	NA	1			1		
1116		NODE_761662_length_262_cov_609.851135	A		1			1			1		1		1		1			1		NP	NP	NP	1			1			1		1		NP	NP	NP			1					
			B		1			1			1		1		1		1			1		NP	NP	NP	1			1			1		1		NP	NP	NP			1					
			C			1		1		NP	NP	NP		1	NP	NP		1	error	error	error	NP	NP	NP		1		1			1		1		NP	NP	NP	NP	NP	NP			1		
			D		1			1				1		1		1		1			1	NP	NP	NP	1			1			1		1		NP	NP	NP			1					

ContinuationTable S3.1. 26 loci (L) sequenced for *Senecio lautus*,their location in the genome (Contig) and in the linkage map (C or chromosome number), and neutrality tests (T) for each one of them for each of the populations in the study. A corresponds to the Fu & Li's D*, B to Fu and Li's F*, C to HKY and D to Tajima's D. For each test we report the significance level (with a number 1 in the column) at which each gene signifincantly deviated from neutrallity: ** p<0.01, * p<0.05 and ns p>0.05. NP stand for non-polymorphic loci, M for loci with missing sequences for the outgroup, and error when the program was not able to conduct the test. Genes where HKA test was performed where judged according to its results. For those genes with a missing outgroup (M), we based our desition Fu and Li's tests and Tajima's D results. Genes in light green are neutrally evolving genes. Genes in dark green are neutrally evolving but presenting recombination events. Genes in white deviated from neutral expectations.

L	C	Contig	Test	A03			A05			D01			D14		D23		D03		D32			D04			H01			H12			H15		H02		H21			H05			I01			I02		
				**	*	ns	**	*	ns	**	*	ns	*	ns	*	ns	*	ns	**	*	ns	**	*	ns	**	*	ns	**	*	ns	*	ns	*	ns	**	*	ns	**	*	ns	**	*	ns	**	*	ns
1125		NODE_198092_length_3695_cov_8.993505	A			1			1			1		1		1	1				1		1			1			1			1		1			1			1			1		
			B			1			1			1		1		1	1				1		1			1			1			1		1			1			1			1		
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1			1			1		1		1	1			1			1			1		1			1		1			1			1			1			1
1126		NODE_2008858_length_1470_cov_13.659184	A			1			1			1		1		1		1			1			1			1			1		NP	NP			1			1			1		
			B			1			1			1		1		1		1			1			1			1			1		NP	NP			1			1			1		
			C					1		error	error	error	NP	NP		1		1			1			1			1			1		NP	NP			1			1			1		
			D			1		1				1		1		1		1			1			1			1			1		NP	NP		1			1			1			1
1170		NODE_2343_length_681_cov_7.656388	A			1			1			1		1		1	NP	NP			1			1			1		NP	NP	NP	NP	NP		1			1			1			1		
			B			1			1			1		1		1	NP	NP			1			1			1		NP	NP	NP	NP	NP		1			1			1			1		
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1			1			1		1		1	NP	NP			1			1			1		NP	NP	NP	NP	NP		1			1			1			1		
1174		NODE_2845_length_2348_cov_10.732965	A		1				1			1		1		1		1		1			1			1			1			1		1			1			1			1		
			B		1				1			1		1		1		1		1			1			1			1			1		1			1			1			1		
			C			1			1			1		1		1		1			1			1			1			1			1			1			1			1		
			D		1				1			1		1		1		1			1			1			1			1			1		1			1			1			1		
1176		NODE_1861_length_664_cov_72.429214	A			1			1			1			1			1			1			1			1			1			1			1			1			1		
			B			1			1			1			1			1			1			1			1			1			1			1			1			1		
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1			1			1		1		1		1			1			1			1			1			1			1			1			1			1
1192	10	NODE_607089_length_1634_cov_8.701959	A			1	NP	NP	NP			1		1		1		1			1			1			1			1			1			1			1			1		
			B			1	NP	NP	NP			1		1		1		1			1			1			1			1			1			1			1			1		
			C			1	NP	NP	NP			1		1		1		1			1			1			1			1			1			1			1			1		
			D			1	NP	NP	NP			1		1		1		1			1			1			1			1			1		1			1			1			1		
1194	6	NODE_123084_length_1557_cov_11.728966	A			1			1			1		1		1		1			1			1			1		NP	NP	NP		1		1			1			1			1		
			B		1				1			1		1		1		1			1			1			1		NP	NP	NP	1			1	1			1			1			
			C			1			1			1		1		1		1			1			1			1		NP	NP	NP		1		1			1			1			1		
			D			1			1			1		1		1		1			1			1			1		NP	NP	NP		1		1	1			1			1			1
1199	19	NODE_1729671_length_1565_cov_10.762300	A			1			1	NP	NP	NP		1		1		1			1			1		1				1			1			1		1						1
			B			1			1	NP	NP	NP		1		1		1			1			1		1		1			1			1			1			1			1
			C			1			1	NP	NP	NP		1		1		1			1			1			1			1			1			1			1			1		
			D			1			1	NP	NP	NP		1		1		1			1			1		1			1			1			1			1			1			1
1204	17	NODE_3424744_length_1169_cov_8.151411	A			1			1			1		1		1		1			1			1			1			1			1		NP	NP	NP			1			1		
			B			1			1			1		1		1		1			1			1			1			1			1		NP	NP	NP			1			1		
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1			1			1		1		1		1			1			1			1			1			1		NP	NP	NP			1			1		
1205	12	NODE_3519433_length_1480_cov_7.394595	A			1			1			1		1		1	NP	NP			1			1			1			1			1			1			1			1			1
			B			1			1			1		1		1	NP	NP			1			1			1			1			1			1			1			1			1
			C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
			D			1			1			1		1		1	NP	NP			1			1			1			1			1			1			1			1			1
1211		G03_NODE_70354_length_249_cov_868.196777	A			1			1		1			1		1		1		1			1			1			1		NP	NP			1			1			1				
			B			1			1		1			1		1		1		1			1			1			1		NP	NP			1			1			1				
			C			1			1			1	NP	NP		1	NP	NP			1			1	NP	NP	NP			1		NP	NP			1			1	NP	NP	NP			1
			D			1			1		1			1		1		1		1			1			1			1		NP	NP		1			1			1			1		
1212		G03_NODE_137343_length_544_cov_677.790466	A	NP	NP	NP			1			1		1		1		1			1			1			1		NP	NP	1			1			1			1	NP	NP	NP		
			B	NP	NP	NP			1			1		1		1		1			1			1			1		NP	NP	1			1			1			1	NP	NP	NP		
			C	NP	NP	NP			1			1		1		1		1			1			1			1		NP	NP		1		1			1			1	NP	NP	NP		
			D	NP	NP	NP			1			1		1		1		1			1			1			1		NP	NP		1		1			1			1	NP	NP	NP		

Table S3.2. Polymorphism analysis for *Senecio lautus* populations. N corresponds to the number of sequences in the analysis, L to the length removing indels, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values). Spaces with an np stand for non polymorphic.

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade		D1	D	10	397	4	5	0.844	0.003	0.00347	-0.5206
Eastern Clade		D3	D	12	397	4	6	0.682	0.00258	0.00323	-0.7411
Eastern Clade		D4	D	10	397	0	1	0	0	0	np
Eastern Clade	1004	H1	H	8	397	3	2	0.429	0.00328	0.00295	0.4577
Eastern Clade		H2	H	14	397	2	3	0.484	0.00171	0.00204	-0.4376
Eastern Clade		H5	H	14	397	6	4	0.396	0.00244	0.00536	-1.9589*
Eastern Clade		I02	inl	8	397	7	4	0.643	0.00479	0.0067	-1.3593
Eastern Clade		D1	D	18	320	2	2	0.111	0.00069	0.00181	-1.5078
Eastern Clade		D3	D	20	320	0	0	0	0	0	np
Eastern Clade		D4	D	22	320	1	2	0.091	0.00028	0.00086	-1.1624
Eastern Clade	1011	H1	H	24	320	2	3	0.42	0.00143	0.0017	-0.3543
Eastern Clade		H2	H	20	320	1	2	0.1	0.00032	0.00091	-1.1644
Eastern Clade		H5	H	22	320	2	2	0.091	0.00056	0.00168	-1.5148
Eastern Clade		I02	inl	20	320	12	8	0.7	0.00616	0.00995	-1.3666
Eastern Clade		D1	D	16	392	3	4	0.642	0.00191	0.00231	-0.4941
Eastern Clade	1014	D3	D	22	392	1	2	0.173	0.00045	0.00071	-0.6411
Eastern Clade		D4	D	22	392	4	4	0.333	0.00127	0.00313	-1.6671

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade		H1	H	22	392	2	3	0.537	0.00158	0.00147	0.1656
Eastern Clade		H2	H	18	392	3	3	0.451	0.00163	0.00242	-0.9027
Eastern Clade	1014	H5	H	22	392	4	4	0.002	0.00293	0.00293	-0.8916
Eastern Clade		I02	inl	20	392	4	4	0.363	0.00124	0.00285	-1.6381
Eastern Clade		D1	D	10	381	0	0	0	0	0	np
Eastern Clade		D3	D	18	381	7	6	0.778	0.00395	0.00534	-0.8869
Eastern Clade		D4	D	16	381	2	3	0.342	0.00099	0.00167	-1.0379
Eastern Clade	1018	H1	H	18	381	8	6	0.562	0.00277	0.00595	-1.8542*
Eastern Clade		H2	H	10	381	2	3	0.378	0.00101	0.00178	-1.4009
Eastern Clade		H5	H	22	381	5	5	0.338	0.00123	0.0037	-1.9873*
Eastern Clade		I02	inl	16	381	15	12	0.958	0.00643	0.01212	-1.8376*
Eastern Clade		D1	D	18	371	13	6	0.562	0.00398	0.00974	-2.2078*
Eastern Clade		D3	D	20	371	4	4	0.284	0.00127	0.00293	-1.6381
Eastern Clade		D4	D	22	371	10	5	0.407	0.00289	0.00739	-2.0756*
Eastern Clade	1021	H1	H	12	371	3	4	0.688	0.00308	0.00296	0.1074
Eastern Clade		H2	H	18	371	10	7	0.739	0.00445	0.00876	-1.7715
Eastern Clade		H5	H	22	371	21	4	0.515	0.00614	0.016	-2.2999**

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade	1021	I02	inl	22	371	9	7	0.771	0.00525	0.0082	-1.2066
Eastern Clade		D1	D	20	418	3	4	0.432	0.00146	0.0023	-0.9752
Eastern Clade		D3	D	22	418	1	2	0.416	0.00105	0.00069	0.8953
Eastern Clade		D4	D	20	418	1	2	0.189	0.00045	0.00067	-0.5916
Eastern Clade	1024	H1	H	22	418	3	4	0.61	0.00166	0.00195	-0.3824
Eastern Clade		H2	H	16	418	2	2	0.125	0.00061	0.00147	-1.498
Eastern Clade		H5	H	22	418	1	2	0.091	0.00023	0.00068	-1.1624
Eastern Clade		I02	inl	22	418	2	3	0.177	0.00046	0.00139	-1.5148
Eastern Clade		D1	D	16	417	3	3	0.425	0.00147	0.00232	-1.0552
Eastern Clade		D3	D	24	417	3	4	0.634	0.00182	0.00193	-0.1331
Eastern Clade		D4	D	22	417	23	4	0.593	0.00585	0.01485	-2.2811**
Eastern Clade	1080	H1	H	22	417	0	0	0	0	0	np
Eastern Clade		H2	H	22	417	3	4	0.606	0.00171	0.00199	-0.3688
Eastern Clade		H5	H	24	417	0	0	0	0	0	np
Eastern Clade		I02	inl	22	417	7	5	0.749	0.0037	0.0046	-0.6318
Eastern Clade		D1	D	8	372	0	0	0	0	0	np
Eastern Clade	1084	D3	D	12	372	1	2	0.303	0.0008	0.00088	-0.1949

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade		D4	D	22	372	7	7	0.541	0.00193	0.00516	-2.0026*
Eastern Clade		H1	H	14	372	6	5	0.505	0.00229	0.00504	-1.9589*
Eastern Clade	1084	H2	H	14	372	6	4	0.582	0.00391	0.00504	-0.809
Eastern Clade		H5	H	22	372	0	0	0	0	0	np
Eastern Clade		I02	inl	12	372	5	6	0.758	0.00261	0.00445	-1.5273
Eastern Clade		D1	D	16	378	14	4	0.642	0.00543	0.01065	-1.9071*
Eastern Clade		D3	D	20	378	4	4	0.616	0.0025	0.00307	-1.0716
Eastern Clade		D4	D	18	378	3	3	0.216	0.00092	0.0024	-1.713
Eastern Clade	1085	H1	H	20	378	3	4	0.726	0.00275	0.0022	0.6705
Eastern Clade		H2	H	18	378	1	2	0.471	0.0012	0.0074	1.1662
Eastern Clade		H5	H	24	378	1	2	0.159	0.00043	0.00073	-0.6811
Eastern Clade		I02	inl	8	378	2	3	0.216	0.00066	0.00172	-1.5078
Eastern Clade		D1	D	18	407	7	7	0.693	0.00353	0.00615	-1.4427
Eastern Clade		D3	D	20	407	9	4	0.284	0.00243	0.00623	-2.0976*
Eastern Clade	1090	D4	D	20	407	12	4	0.284	0.00299	0.00844	-2.3162**
Eastern Clade		H1	H	22	407	2	2	0.091	0.00052	0.00156	-1.5148
Eastern Clade		H2	H	18	407	17	4	0.314	0.00645	0.01347	-2.0025*

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade		H5	H	22	407	0	0	0	0	0	np
Eastern Clade	1090	I02	inl	8	407	1	2	0.536	0.00144	0.00103	1.1665
Eastern Clade		D1	D	12	395	11	7	0.773	0.00816	0.00944	-0.5609
Eastern Clade		D3	D	10	395	6	2	0.356	0.00572	0.00569	0.0242
Eastern Clade		D4	D	8	395	6	4	0.643	0.00469	0.00584	-0.9204
Eastern Clade	1091	H1	H	18	395	3	4	0.643	0.00192	0.00296	-1.4475
Eastern Clade		H2	H	12	395	13	5	0.667	0.01053	0.01104	-0.1956
Eastern Clade		H5	H	12	395	7	5	0.742	0.00943	0.00653	1.7398
Eastern Clade		I02	inl	8	395	4	3	0.607	0.00497	0.00438	0.5862
Eastern Clade		D1	D	12	393	3	3	0.439	0.0022	0.00282	-0.7287
Eastern Clade		D3	D	16	393	7	4	0.525	0.00428	0.00593	-0.9764
Eastern Clade		D4	D	10	393	9	5	0.756	0.00642	0.00901	-1.259
Eastern Clade	1096	H1	H	6	393	9	6	1	0.01001	0.01003	-0.0125
Eastern Clade		H2	H	8	393	4	4	0.821	0.00476	0.00396	0.8892
Eastern Clade		H5	H	12	393	5	5	0.788	0.00418	0.00481	-0.4816
Eastern Clade		I02	inl	12	393	10	7	0.879	0.00895	0.01093	-0.7445
Eastern Clade	1098	D1	D	20	377	8	8	0.647	0.00283	0.00648	-2.1074*

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade		D3	D	22	377	4	3	0.177	0.00094	0.00284	-1.8776*
Eastern Clade		D4	D	22	377	5	7	0.688	0.00242	0.00364	-0.9929
Eastern Clade		H1	H	22	377	2	2	0.091	0.00046	0.0014	-1.5148
Eastern Clade	1098	H2	H	22	377	7	5	0.338	0.00202	0.00541	-2.0025*
Eastern Clade		H5	H	2	377	1	2	1	0.00255	0.00255	np
Eastern Clade		I02	inl	22	377	8	7	0.541	0.00375	0.00844	-1.8258*
Eastern Clade		D1	D	20	278	2	3	0.195	0.00072	0.00203	-1.5128
Eastern Clade		D3	D	18	278	2	3	0.216	0.0008	0.00208	-1.5078
Eastern Clade		D4	D	16	278	0	0	0	0	0	np
Eastern Clade	1116	H1	H	24	278	11	2	0.083	0.00332	0.01067	-2.3329**
Eastern Clade		H2	H	22	278	0	0	0	0	0	np
Eastern Clade		H5	H	20	278	1	2	0.111	0.00044	0.00116	-1.8679*
Eastern Clade		I02	inl	20	278	3	4	0.284	0.00108	0.00303	-1.7233
Eastern Clade		D1	D	8	437	3	4	0.75	0.00212	0.00265	-0.8125
Eastern Clade		D3	D	16	437	5	4	0.35	0.00155	0.00375	-1.9286*
Eastern Clade	1125	D4	D	10	437	7	3	0.378	0.00324	0.00573	-1.8391*
Eastern Clade		H1	H	14	437	21	8	0.868	0.01122	0.01751	-1.5215

Continuation Table S.3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade		H2	H	12	437	15	6	0.758	0.00711	0.01214	-1.7789*
Eastern Clade	1125	H5	H	8	437	1	2	0.429	0.00101	0.00091	0.3335
Eastern Clade		I02	inl	12	437	5	6	0.778	0.00328	0.00408	-0.7832
Eastern Clade		D1	D	8	408	2	2	0.25	0.00119	0.00183	-1.3101
Eastern Clade		D3	D	12	408	1	2	0.167	0.00041	0.00082	-1.1405
Eastern Clade		D4	D	10	408	7	7	0.933	0.00419	0.00606	-1.3066
Eastern Clade	1126	H1	H	16	408	3	3	0.242	0.001	0.00242	-1.6965
Eastern Clade		H2	H	6	408	0	0	0	0	0	np
Eastern Clade		H5	H	14	408	4	3	0.275	0.00159	0.0035	-1.7976*
Eastern Clade		I02	inl	12	408	4	5	0.576	0.00203	0.00335	-1.3848
Eastern Clade		D1	D	12	403	4	4	0.561	0.00266	0.00327	-0.6606
Eastern Clade		D3	D	12	403	0	1	0	0	0	np
Eastern Clade		D4	D	8	403	1	2	0.429	0.00105	0.00094	0.3335
Eastern Clade	1170	H1	H	14	403	1	2	0.363	0.00106	0.00092	0.3244
Eastern Clade		H2	H	14	403	2	3	0.385	0.00123	0.0019	-0.9592
Eastern Clade		H5	H	8	403	2	3	0.607	0.00167	0.0019	-0.4479
Eastern Clade		I02	inl	10	403	7	5	0.667	0.00361	0.00638	-1.8391*

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade		D1	D	16	362	6	5	0.608	0.00316	0.00483	-1.1829
Eastern Clade		D3	D	16	362	3	3	0.342	0.00214	0.0025	-0.414
Eastern Clade		D4	D	18	362	3	3	0.386	0.00136	0.0023	-1.1313
Eastern Clade	1174	H1	H	20	362	5	5	0.368	0.00163	0.00389	-1.78
Eastern Clade		H2	H	16	362	2	3	0.242	0.00069	0.00166	-1.498
Eastern Clade		H5	H	20	362	13	6	0.658	0.00499	0.01018	-1.8507*
Eastern Clade		I02	inl	16	362	3	4	0.65	0.00216	0.00252	-0.414
Eastern Clade		D1	D	12	398	2	3	0.318	0.00084	0.00166	-1.179
Eastern Clade		D3	D	12	398	4	4	0.561	0.00142	0.00234	-1.3848
Eastern Clade		D4	D	12	398	3	4	0.455	0.00135	0.00174	-0.7287
Eastern Clade	1176	H1	H	12	398	1	2	0.303	0.00076	0.00083	-0.1949
Eastern Clade		H2	H	12	398	2	3	0.318	0.00084	0.00166	-1.4514
Eastern Clade		H5	H	12	398	1	2	0.303	0.00076	0.00083	-0.1949
Eastern Clade		I02	inl	10	398	4	4	0.533	0.00244	0.00353	-1.7915*
Eastern Clade		D1	D	4	399	2	3	0.833	0.00287	0.00268	0.5916
Eastern Clade	1192	D3	D	20	399	1	2	0.189	0.00048	0.00071	-0.5916
Eastern Clade		D4	D	12	399	9	5	0.576	0.00536	0.00732	-1.0875

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade		H1	H	10	399	5	5	0.784	0.00431	0.00353	0.6767
Eastern Clade		H2	H	18	399	1	2	0.111	0.00029	0.00076	-1.1647
Eastern Clade	1192	H5	H	18	399	5	3	0.503	0.00548	0.0036	1.6592
Eastern Clade		I02	inl	16	399	2	3	0.242	0.00063	0.00151	-1.498
Eastern Clade		D1	D	20	372	1	2	0.442	0.00119	0.00076	1.0259
Eastern Clade		D3	D	22	372	1	2	0.485	0.0013	0.00074	1.3343
Eastern Clade		D4	D	22	372	1	2	0.173	0.00047	0.00074	-0.6411
Eastern Clade	1194	H1	H	24	372	1	2	0.159	0.00043	0.00072	-0.6811
Eastern Clade		H2	H	22	372	1	2	0.091	0.00025	0.00075	-1.162
Eastern Clade		H5	H	24	372	17	3	0.453	0.0046	0.01214	-2.2270**
Eastern Clade		I02	inl	8	372	10	4	0.26	0.0028	0.00775	-2.1778**
Eastern Clade		D1	D	12	386	0	0	0	0	0	np
Eastern Clade		D3	D	18	386	6	7	0.562	0.00331	0.00574	-1.7151
Eastern Clade		D4	D	12	386	9	5	0.742	0.00611	0.00903	-1.6104
Eastern Clade	1199	H1	H	14	386	11	4	0.626	0.0051	0.00896	-1.9065*
Eastern Clade		H2	H	14	386	3	5	0.593	0.00191	0.00225	-1.1932
Eastern Clade		H5	H	18	386	9	9	0.889	0.00695	0.00721	-0.4762

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade	1199	I02	inl	14	386	27	6	0.769	0.0166	0.02295	-1.2922
Eastern Clade		D1	D	12	410	2	3	0.439	0.00116	0.00164	-0.8497
Eastern Clade		D3	D	14	410	2	3	0.275	0.00072	0.00158	-1.4807
Eastern Clade		D4	D	12	410	1	2	0.53	0.00129	0.00081	1.3811
Eastern Clade	1204	H1	H	10	410	2	2	0.356	0.00173	0.00172	0.0189
Eastern Clade		H2	H	14	410	1	2	0.495	0.00132	0.00084	1.2122
Eastern Clade		H5	H	22	410	3	4	0.455	0.00142	0.00212	-0.8586
Eastern Clade		I02	inl	8	410	3	6	0.893	0.00551	0.00851	-1.7232*
Eastern Clade		D1	D	10	365	2	2	0.2	0.00108	0.00192	-1.4009
Eastern Clade		D3	D	12	365	0	0	0	0	0	np
Eastern Clade		D4	D	14	365	2	3	0.385	0.00115	0.00178	-0.9592
Eastern Clade	1205	H1	H	12	365	4	3	0.439	0.0022	0.00363	-1.3848
Eastern Clade		H2	H	10	365	2	3	0.378	0.00114	0.00201	-1.4009
Eastern Clade		H5	H	18	365	3	4	0.314	0.00091	0.00238	-1.713
Eastern Clade		I02	inl	14	365	3	4	0.495	0.00211	0.00259	-0.5651
Eastern Clade		D1	D	14	291	3	4	0.239	0.00112	0.00273	-1.8931*
Eastern Clade	1211	D3	D	22	291	2	3	0.385	0.00139	0.00189	-0.6031

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Eastern Clade		D4	D	24	291	6	7	0.605	0.00294	0.00545	-1.3944
Eastern Clade		H1	H	22	291	1	2	0.091	0.00031	0.00094	-1.1624
Eastern Clade	1211	H2	H	24	291	8	4	0.431	0.00476	0.00779	-1.2495
Eastern Clade		H5	H	22	291	3	4	0.333	0.00148	0.00279	-1.2124
Eastern Clade		I02	inl	22	291	4	6	0.411	0.00191	0.00465	-1.9966*
Eastern Clade		D1	D	24	428	2	3	0.236	0.00057	0.00125	-1.2023
Eastern Clade		D3	D	24	428	2	3	0.236	0.00073	0.00126	-0.9196
Eastern Clade		D4	D	24	428	2	3	0.163	0.00039	0.00127	-1.5147
Eastern Clade	1212	H1	H	24	428	4	5	0.591	0.00224	0.0025	-0.2834
Eastern Clade		H2	H	24	428	3	3	0.163	0.00059	0.0019	-1.7325
Eastern Clade		H5	H	24	428	4	5	0.529	0.00157	0.0025	-1.0235
Eastern Clade		I02	inl	22	428	0	0	0	0	0	np

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		A03	alp	8	397	5	4	0.643	0.00306	0.00473	-1.5952
Southern Clade		A05	alp	10	397	3	4	0.644	0.0019	0.00267	-1.0345
Southern Clade		D14	D	14	397	3	4	0.714	0.0251	0.00254	-0.0302
Southern Clade		D23	D	10	397	4	4	0.533	0.002	0.00353	-1.6671
Southern Clade		D32	D	14	397	6	4	0.396	0.00216	0.00476	-1.9589*
Southern Clade	1004	H12	H	14	397	2	3	0.385	0.00132	0.00164	-0.5325
Southern Clade		H15	H	8	397	11	6	0.893	0.01018	0.01042	-0.1172
Southern Clade		H21	H	12	397	1	2	0.303	0.00075	0.00082	-0.1949
Southern Clade		I01	inl	6	397	3	2	0.333	0.00254	0.00334	-1.2331
Southern Clade		A03	alp	20	320	1	2	0.189	0.00052	0.00078	-0.5916
Southern Clade		A05	alp	22	320	0	1	0	0	0	np
Southern Clade		D14	D	24	320	0	0	0	0	0	np
Southern Clade	1011	D23	D	16	320	6	8	0.825	0.00389	0.0057	-1.0882
Southern Clade		D32	D	22	320	8	8	0.779	0.00518	0.00625	-0.8834
Southern Clade		H12	H	24	320	5	7	0.558	0.00198	0.00373	-1.6876

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		H15	H	18	320	2	3	0.216	0.00062	0.00163	-1.5078
Southern Clade		H21	H	24	320	1	2	0.464	0.00144	0.00083	1.2318
Southern Clade	1011	I01	inl	22	320	6	5	0.472	0.00225	0.00521	-1.7623
Southern Clade		A03	alp	18	392	2	3	0.307	0.00081	0.00148	-1.0963
Southern Clade		A05	alp	22	392	2	3	0.177	0.00046	0.0014	-1.5148
Southern Clade		D14	D	24	392	3	3	0.163	0.00065	0.0021	-1.7326
Southern Clade		D23	D	22	392	8	5	0.649	0.00295	0.00567	-1.5732
Southern Clade		D32	D	22	392	4	5	0.407	0.00126	0.00309	-1.6671
Southern Clade	1014	H12	H	22	392	4	5	0.338	0.00116	0.00282	-1.9966*
Southern Clade		H15	H	22	392	1	2	0.091	0.00024	0.00072	-1.1624
Southern Clade		H21	H	24	392	6	7	0.446	0.00221	0.0045	-1.524
Southern Clade		I01	inl	22	392	10	10	0.792	0.00338	0.00707	-1.7796
Southern Clade		I02	inl	20	392	4	4	0.363	0.00124	0.00285	-1.6381
Southern Clade		A03	alp	12	381	4	6	0.848	0.00338	0.00343	-0.0572
Southern Clade		A05	alp	18	381	0	0	0	0	0	np
Southern Clade	1018	D14	D	22	381	3	4	0.333	0.00091	0.0021	-1.4709
Southern Clade		D23	D	20	381	5	4	0.363	0.0015	0.00358	-1.78

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		D32	D	18	381	14	8	0.641	0.00459	0.01066	-2.1435**
Southern Clade		H12	H	20	381	3	5	0.442	0.00146	0.00221	-0.9088
Southern Clade	1018	H15	H	4	381	6	3	0.833	0.0075	0.00818	-0.8086
Southern Clade		H21	H	20	381	13	10	0.863	0.00568	0.0094	-1.4339
Southern Clade		I01	inl	6	381	7	5	0.933	0.00883	0.00766	0.8878
Southern Clade		A03	alp	18	371	7	3	0.216	0.00274	0.00717	-2.0965*
Southern Clade		A05	alp	20	371	17	9	0.847	0.00808	0.01484	-1.7028
Southern Clade		D14	D	24	371	23	7	0.699	0.00669	0.01555	-2.1052*
Southern Clade		D23	D	2	371	0	0	0	0	0	np
Southern Clade	1021	D32	D	22	371	5	6	0.589	0.00208	0.0035	-1.2086
Southern Clade		H12	H	22	371	9	9	0.658	0.00377	0.00627	-1.3359
Southern Clade		H15	H	20	371	8	7	0.774	0.0046	0.00684	-1.1273
Southern Clade		H21	H	16	371	18	9	0.892	0.01655	0.01511	0.1509
Southern Clade		I01	inl	20	371	12	7	0.521	0.00416	0.00872	-1.8749*
Southern Clade		A03	alp	22	418	3	3	0.177	0.00066	0.001199	-1.7294
Southern Clade	1024	A05	alp	22	418	3	3	0.177	0.00089	0.00206	-1.4709
Southern Clade		D14	D	24	418	0	0	0	0	0	np

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		D23	D	24	418	3	4	0.37	0.00124	0.00218	-1.0859
Southern Clade		D32	D	22	418	9	7	0.481	0.00203	0.006611	-2.2410**
Southern Clade		H12	H	24	418	5	4	0.37	0.00135	0.00321	-1.6824
Southern Clade	1024	H15	H	22	418	2	3	0.329	0.0009	0.00142	-0.8355
Southern Clade		H21	H	24	418	7	7	0.652	0.00269	0.00471	-1.3373
Southern Clade		I01	inl	22	418	4	4	0.333	0.00106	0.00261	-1.6671
Southern Clade		A03	alp	22	417	4	4	0.524	0.00211	0.00261	-0.537
Southern Clade		A05	alp	22	417	7	7	0.541	0.00194	0.00466	-1.8655*
Southern Clade		D14	D	24	417	2	3	0.562	0.00142	0.00126	0.2853
Southern Clade		D23	D	22	417	4	5	0.753	0.0024	0.00264	-0.249
Southern Clade	1080	D32	D	22	417	4	4	0.333	0.00124	0.00261	-1.4788
Southern Clade		H12	H	24	417	3	4	0.471	0.00133	0.00192	-0.7683
Southern Clade		H15	H	22	417	2	3	0.55	0.00145	0.0013	0.2729
Southern Clade		H21	H	24	417	6	6	0.681	0.00258	0.00389	-1.0193
Southern Clade		I01	inl	22	417	16	10	0.879	0.00594	0.01097	-1.9080*
Southern Clade		A03	alp	12	372	10	8	0.924	0.00627	0.00836	-1.0269
Southern Clade	1084	A05	alp	16	372	5	4	0.525	0.00286	0.00407	-0.9809

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		D14	D	12	372	1	2	0.303	0.00078	0.00085	-0.1949
Southern Clade		D23	D	10	372	2	3	0.378	0.00105	0.00186	-1.4009
Southern Clade		D32	D	16	372	9	7	0.692	0.00489	0.00723	-1.1902
Southern Clade	1084	H12	H	14	372	1	2	0.143	0.00037	0.00082	-1.1552
Southern Clade		H15	H	10	372	2	3	0.622	0.00187	0.00186	0.0189
Southern Clade		H21	H	14	372	4	5	0.593	0.00237	0.00327	-0.9054
Southern Clade		I01	inl	16	372	3	4	0.442	0.00129	0.00241	-1.3492
Southern Clade		A03	alp	8	378	1	2	0.429	0.00113	0.00102	0.3335
Southern Clade		A05	alp	12	378	16	3	0.318	0.00684	0.01359	-2.1445**
Southern Clade		D14	D	14	378	3	3	0.385	0.00145	0.0025	-1.2783
Southern Clade		D23	D	16	378	5	2	0.125	0.00162	0.0039	-1.9286*
Southern Clade	1085	D32	D	22	378	4	2	0.091	0.00097	0.00293	-1.8776*
Southern Clade		H12	H	20	378	3	4	0.284	0.0008	0.00225	-1.7233
Southern Clade		H15	H	16	378	2	3	0.242	0.00073	0.00176	-1.498
Southern Clade		H21	H	12	378	2	3	0.545	0.00163	0.00179	-0.2481
Southern Clade		I01	inl	18	378	18	6	0.604	0.00723	0.01455	-2.0850*
Southern Clade	1090	A03	alp	18	407	5	5	0.601	0.002	0.00359	-1.4006

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		A05	alp	12	407	3	3	0.318	0.00126	0.0025	-1.6293
Southern Clade		D14	D	14	407	3	2	0.143	0.00104	0.00228	-1.6705
Southern Clade		D23	D	22	407	4	4	0.26	0.00105	0.00316	-1.8776*
Southern Clade		D32	D	18	407	3	3	0.216	0.00139	0.00229	-1.0897
Southern Clade	1090	H12	H	12	407	5	4	0.742	0.00394	0.00418	-0.2117
Southern Clade		H15	H	16	407	4	3	0.242	0.00136	0.00328	-1.8309*
Southern Clade		H21	H	20	407	8	6	0.447	0.00197	0.00554	-2.1746**
Southern Clade		I01	inl	4	407	8	6	0.929	0.0056	0.00793	-1.4213
Southern Clade		A03	alp	12	395	8	9	0.939	0.00535	0.00678	-0.8415
Southern Clade		A05	alp	14	395	9	8	0.868	0.0063	0.00811	-0.8574
Southern Clade		D14	D	12	395	4	4	0.712	0.00414	0.00335	0.8279
Southern Clade		D23	D	12	395	3	4	0.652	0.00195	0.00251	-0.7287
Southern Clade	1091	D32	D	12	395	2	2	0.167	0.00101	0.00201	-1.4514
Southern Clade		H12	H	14	395	5	4	0.396	0.00191	0.00422	-1.8893*
Southern Clade		H15	H	10	395	5	5	0.8	0.0052	0.00454	0.5781
Southern Clade		H21	H	16	395	7	6	0.542	0.00364	0.00588	-1.3383
Southern Clade		I01	inl	14	395	10	8	0.923	0.00757	0.00888	-0.5736

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		A03	alp	8	393	1	2	0.536	0.00134	0.00096	1.1665
Southern Clade		A05	alp	14	393	1	2	0.143	0.00041	0.00091	-1.1552
Southern Clade		D14	D	14	393	0	0	0	0	0	np
Southern Clade		D23	D	6	393	0	0	0	0	0	np
Southern Clade	1096	D32	D	14	393	2	3	0.385	0.00123	0.0019	-0.9592
Southern Clade		H12	H	6	393	2	3	0.8	0.0027	0.00222	1.0319
Southern Clade		H15	H	14	393	0	0	0	0	0	np
Southern Clade		H21	H	12	393	3	4	0.561	0.00227	0.00304	-0.8288
Southern Clade		I01	inl	6	393	3	2	0.333	0.00272	0.00358	-1.2331
Southern Clade		A03	alp	22	377	6	5	0.468	0.00183	0.0443	-1.7786
Southern Clade		A05	alp	22	377	6	6	0.476	0.00182	0.00421	-1.7623
Southern Clade		D14	D	24	377	7	8	0.757	0.0032	0.00487	-1.3635
Southern Clade		D23	D	24	377	11	7	0.605	0.0039	0.00796	-1.8912*
Southern Clade	1098	D32	D	22	377	5	6	0.58	0.001184	0.00345	-1.3869
Southern Clade		H12	H	24	377	3	3	0.163	0.00064	0.00205	-1.7325
Southern Clade		H15	H	24	377	3	5	0.591	0.0017	0.00205	-1.042
Southern Clade		H21	H	24	377	9	7	0.504	0.00217	0.00633	-2.1542*

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade	1098	I01	inl	22	377	13	11	0.758	0.00342	0.00912	-2.2134**
Southern Clade		A03	alp	18	278	5	4	0.314	0.002	0.00523	-1.9556*
Southern Clade		A05	alp	18	278	1	2	0.111	0.00044	0.00116	-1.1647
Southern Clade		D14	D	22	278	1	2	0.091	0.00033	0.00099	-1.1624
Southern Clade		D23	D	22	278	1	2	0.091	0.00032	0.00098	-1.1624
Southern Clade	1116	D32	D	16	278	1	2	0.125	0.00045	0.00108	-1.1622
Southern Clade		H12	H	24	278	2	3	0.163	0.0006	0.00193	-1.5147
Southern Clade		H15	H	20	278	3	3	0.195	0.00109	0.00306	-1.7233
Southern Clade		H21	H	24	278	1	2	0.083	0.00031	0.00098	-1.1593
Southern Clade		I01	inl	22	278	0	0	0	0	0	np
Southern Clade		A03	alp	10	437	5	5	0.667	0.00322	0.0045	-1.1361
Southern Clade		A05	alp	12	437	1	2	0.167	0.00051	0.00101	-1.1405
Southern Clade		D14	D	12	437	17	5	0.667	0.00825	0.01321	-1.6297
Southern Clade	1125	D23	D	14	437	13	8	0.89	0.00661	0.00957	-1.4518
Southern Clade		D32	D	14	437	16	9	0.835	0.00625	0.01187	-1.9431*
Southern Clade		H12	H	12	437	18	6	0.682	0.00754	0.01367	-2.0894**
Southern Clade		H15	H	12	437	4	4	0.636	0.00271	0.00312	-0.4595

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		H21	H	12	437	4	4	0.533	0.00223	0.0033	-1.2447
Southern Clade	1125	I01	inl	12	437	5	5	0.867	0.00429	0.00468	-0.3294
Southern Clade		A03	alp	10	408	3	3	0.378	0.00201	0.00282	-1.0345
Southern Clade		A05	alp	10	408	5	2	0.2	0.00244	0.00432	-1.7411*
Southern Clade		D14	D	16	408	1	2	0.125	0.00031	0.00075	-1.1622
Southern Clade		D23	D	18	408	5	6	0.562	0.00234	0.00375	-1.1871
Southern Clade		D32	D	14	408	15	2	0.143	0.00529	0.01165	-2.2252**
Southern Clade	1126	H12	H	10	408	3	3	0.378	0.00143	0.00252	-1.5622
Southern Clade		H15	H	10	408	1	2	0.2	0.00049	0.00087	-1.1112
Southern Clade		H21	H	16	408	7	6	0.542	0.00248	0.00533	-1.8811*
Southern Clade		I01	inl	10	408	1	2	0.356	0.00097	0.00096	0.015
Southern Clade		A03	alp	12	403	5	4	0.773	0.00454	0.00407	0.4292
Southern Clade		A05	alp	10	403	1	2	0.2	0.00049	0.00087	-1.1117
Southern Clade		D14	D	16	403	1	2	0.143	0.00039	0.00085	-1.1552
Southern Clade	1170	D23	D	12	403	1	2	0.167	0.00055	0.00109	-1.1405
Southern Clade		D32	D	12	403	7	4	0.652	0.00547	0.00639	-0.5624
Southern Clade		H12	H	12	403	0	0	0	0	0	np

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		H15	H	6	403	0	0	0	0	0	np
Southern Clade	1170	H21	H	18	403	1	2	0.111	0.00034	0.00089	-1.1647
Southern Clade		I01	inl	10	403	6	5	0.667	0.00352	0.00542	-1.8772*
Southern Clade		A03	alp	20	362	13	7	0.689	0.0047	0.01009	-1.9393*
Southern Clade		A05	alp	20	362	2	3	0.195	0.00131	0.00255	-1.1407
Southern Clade		D14	D	16	362	8	4	0.717	0.00509	0.00679	-0.9011
Southern Clade		D23	D	20	362	5	7	0.737	0.00314	0.00504	-0.7059
Southern Clade	1174	D32	D	20	362	7	6	0.516	0.00236	0.00529	-1.8233*
Southern Clade		H12	H	18	362	6	7	0.739	0.00352	0.00463	-0.7882
Southern Clade		H15	H	20	362	6	6	0.632	0.00323	0.00539	-1.2727
Southern Clade		H21	H	20	362	9	9	0.795	0.00462	0.00649	-0.9917
Southern Clade		I01	inl	22	362	6	4	0.561	0.00338	0.00591	-1.6325
Southern Clade		A03	alp	12	398	2	3	0.318	0.00084	0.00167	-1.4514
Southern Clade		A05	alp	12	398	1	2	0.167	0.00029	0.00057	-1.1405
Southern Clade	1176	D14	D	12	398	6	6	0.758	0.00332	0.00349	-0.1783
Southern Clade		D23	D	12	398	9	4	0.682	0.00538	0.00543	-0.035
Southern Clade		D32	D	12	398	2	2	0.303	0.00107	0.00117	-0.2481

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		H12	H	12	398	1	2	0.303	0.00053	0.00058	-0.1949
Southern Clade		H15	H	12	398	1	2	0.303	0.00053	0.00058	-0.1949
Southern Clade	1176	H21	H	12	398	1	2	0.545	0.00094	0.00057	1.4862
Southern Clade		I01	inl	10	398	1	2	0.2	0.00051	0.00089	-1.1117
Southern Clade		A03	alp	8	399	2	3	0.714	0.00214	0.00193	0.4142
Southern Clade		A05	alp	12	399	0	0	0	0	0	np
Southern Clade		D14	D	22	399	5	5	0.407	0.00154	0.00346	-1.6495
Southern Clade		D23	D	16	399	3	3	0.433	0.00172	0.00223	-0.6544
Southern Clade	1192	D32	D	18	399	5	6	0.562	0.0022	0.00376	-1.3152
Southern Clade		H12	H	14	399	2	2	0.264	0.00135	0.0016	-0.4376
Southern Clade		H15	H	14	399	4	4	0.495	0.00259	0.00349	-0.8479
Southern Clade		H21	H	22	399	7	6	0.411	0.00201	0.00482	-1.8655*
Southern Clade		I01	inl	20	399	9	7	0.784	0.00449	0.00657	-1.0916
Southern Clade		A03	alp	22	372	3	3	0.177	0.00075	0.00225	-1.7294
Southern Clade		A05	alp	22	372	5	5	0.519	0.002	0.00406	-1.5088
Southern Clade	1194	D14	D	24	372	1	2	0.522	0.00151	0.00078	1.5961
Southern Clade		D23	D	24	372	4	4	0.308	0.00138	0.00304	-1.4953

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		D32	D	20	372	2	2	0.1	0.00061	0.00171	-1.5128
Southern Clade		H12	H	24	372	0	0	0	0	0	np
Southern Clade	1194	H15	H	22	372	3	3	0.177	0.00076	0.00231	-1.7294
Southern Clade		H21	H	24	372	11	5	0.377	0.00265	0.00788	-2.2454**
Southern Clade		I01	inl	22	372	3	4	0.26	0.00075	0.00226	-1.7294
Southern Clade		A03	alp	16	386	26	8	0.867	0.01704	0.0236	-1.3572
Southern Clade		A05	alp	16	386	3	4	0.717	0.00411	0.00402	-0.7279
Southern Clade		D14	D	14	386	14	10	0.945	0.01268	0.01572	-1.3861
Southern Clade		D23	D	14	386	9	8	0.857	0.00734	0.0108	-1.7501
Southern Clade	1199	D32	D	16	386	23	6	0.617	0.02738	0.02539	0.3189
Southern Clade		H12	H	16	386	21	7	0.625	0.01069	0.02293	-2.1559**
Southern Clade		H15	H	10	386	12	5	0.822	0.00963	0.01202	-1.6525
Southern Clade		H21	H	16	386	3	3	0.242	0.00128	0.0031	-1.6965
Southern Clade		I01	inl	16	386	9	9	0.892	0.00858	0.0102	-0.5809
Southern Clade		A03	alp	12	410	1	2	0.167	0.00043	0.00085	-1.1405
Southern Clade	1204	A05	alp	12	410	5	6	0.682	0.00248	0.00423	-1.5273
Southern Clade		D14	D	12	410	1	2	0.303	0.00078	0.00086	-0.1949

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		D23	D	12	410	0	0	0	0	0	-0.8497
Southern Clade		D32	D	12	410	2	3	0.439	0.00115	0.00162	-0.8497
Southern Clade		H12	H	12	410	2	3	0.318	0.00082	0.00164	-1.4514
Southern Clade	1204	H15	H	8	410	2	2	0.25	0.00122	0.00188	-1.3101
Southern Clade		H21	H	12	410	0	0	0	0	0	np
Southern Clade		I01	inl	8	410	2	3	0.607	0.0017	0.00194	-0.4479
Southern Clade		A03	alp	12	365	4	5	0.576	0.00189	0.00375	-1.7469
Southern Clade		A05	alp	16	365	11	7	0.75	0.00498	0.00913	-1.7172
Southern Clade		D14	D	18	365	5	3	0.216	0.00224	0.00384	-1.3152
Southern Clade		D23	D	8	365	2	3	0.607	0.00185	0.00211	-0.4479
Southern Clade	1205	D32	D	8	365	2	2	0.25	0.00131	0.00202	-1.3101
Southern Clade		H12	H	10	365	6	4	0.644	0.00412	0.00578	-1.1895
Southern Clade		H15	H	12	365	4	4	0.455	0.0019	0.00377	-1.7469
Southern Clade		H21	H	12	365	2	3	0.318	0.00093	0.00184	-1.4514
Southern Clade		I01	inl	14	365	6	6	0.604	0.00237	0.00521	-1.9589*
Southern Clade		A03	alp	18	291	8	7	0.569	0.00527	0.00791	-1.1608
Southern Clade		A05	alp	22	291	4	4	0.333	0.00155	0.00381	-1.6671

Continuation Table S3.2. Polymorphism analysis for populations and genes in the study. N corresponds to the number of sequences in the analysis, L to the length removing indels and missing data sites, S to the number of segregating sites, H the haplotype number, Hd haplotype diversity, π to the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta to the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D for the estimates of Tajima's D (* accompany significant values).

Region	**Loci**	**Population**	**ecotype**	**N**	**L**	**S**	**H**	**Hd**	**π**	**Theta**	**D**
Southern Clade		D14	D	24	291	3	4	0.308	0.00162	0.00277	-1.0406
Southern Clade		D23	D	22	291	2	3	0.177	0.00063	0.00191	-1.5148
Southern Clade	1211	D32	D	24	291	4	4	0.239	0.00116	0.00373	-1.8838*
Southern Clade		H12	H	24	291	1	2	0.083	0.00029	0.00093	-1.1593
Southern Clade		H15	H	24	291	0	0	0	0	0	np
Southern Clade		H21	H	22	291	0	0	0	0	0	np
Southern Clade		I01	inl	22	291	2	3	0.177	0.00064	0.00194	-1.5148
Southern Clade		A03	alp	22	428	0	0	0	0	0	np
Southern Clade		A05	alp	22	428	1	2	0.091	0.00021	0.00064	-1.1624
Southern Clade	1212	D14	D	24	428	2	3	0.163	0.00039	0.00126	-1.5147
Southern Clade		D23	D	24	428	5	5	0.377	0.00136	0.00323	-1.6824
Southern Clade		D32	D	24	428	2	3	0.163	0.0004	0.00129	-1.5147
Southern Clade		H12	H	24	428	2	3	0.236	0.00092	0.00127	-0.6072
Southern Clade	1212	H15	H	24	428	0	0	0	0	0	np
Southern Clade		H21	H	24	428	3	4	0.373	0.00096	0.00191	-1.2561
Southern Clade		I01	inl	22	428	1	2	0.173	0.00041	0.00065	-0.6411

Table S3.3. Polymorphism analysis for ecotypes and genes in the study. Mean and Standard errors are reported for: N or the number of sequences in the analysis, L or the length removing indels and missing data sites, S or the number of segregating sites, H or haplotype number, Hd or haplotype diversity, π or the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta or the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D or estimated of Tajima's D (* accompany significant values).

Ecotype	loci	L	n		S		H		Hd		π		Theta		D	
			Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
	1004	397	9.000	1.000	4.000	1.000	4.000	0.000	0.644	0.000	0.002	0.001	0.004	0.001	-1.315	0.280
	1011	320	21.000	1.000	0.500	0.500	1.500	0.500	0.095	0.095	0.000	0.000	0.000	0.000	-0.592	
	1014	392	20.000	2.000	2.000	0.000	3.000	0.000	0.242	0.065	0.001	0.000	0.001	0.000	-1.306	0.209
	1018	381	15.000	3.000	2.000	2.000	3.000	3.000	0.424	0.424	0.002	0.002	0.002	0.002	-0.057	
	1021	371	19.000	1.000	12.000	5.000	6.000	3.000	0.532	0.316	0.005	0.003	0.011	0.004	-1.900	0.197
	1024	418	22.000	0.000	3.000	0.000	3.000	0.000	0.177	0.000	0.001	0.000	0.002	0.000	-1.600	0.129
	1080	417	22.000	0.000	5.500	1.500	5.500	1.500	0.533	0.009	0.002	0.000	0.004	0.001	-1.201	0.664
	1084	372	14.000	2.000	7.500	2.500	6.000	2.000	0.725	0.200	0.005	0.002	0.006	0.002	-1.004	0.023
Alpine	1085	378	10.000	2.000	8.500	7.500	2.500	0.500	0.374	0.056	0.004	0.003	0.007	0.006	-0.906	1.239
	1090	407	15.000	3.000	4.000	1.000	4.000	1.000	0.460	0.142	0.002	0.000	0.003	0.001	-1.515	0.114
	1091	395	13.000	1.000	8.500	0.500	8.500	0.500	0.904	0.036	0.006	0.000	0.007	0.001	-0.849	0.008
	1096	393	11.000	3.000	1.000	0.000	2.000	0.000	0.340	0.197	0.001	0.000	0.001	0.000	0.006	1.161
	1098	377	22.000	0.000	6.000	0.000	5.500	0.500	0.472	0.004	0.002	0.000	0.024	0.020	-1.770	0.008
	1116	278	18.000	0.000	3.000	2.000	3.000	1.000	0.213	0.102	0.001	0.001	0.003	0.002	-1.560	0.395
	1125	437	11.000	1.000	3.000	2.000	3.500	1.500	0.417	0.250	0.002	0.001	0.003	0.002	-1.138	0.002
	1126	408	10.000	0.000	4.000	1.000	2.500	0.500	0.289	0.089	0.002	0.000	0.004	0.001	-1.388	0.353
	1170	403	11.000	1.000	3.000	2.000	3.000	1.000	0.487	0.287	0.003	0.002	0.002	0.002	-0.341	0.770

Continuation Table S3.3. Polymorphism analysis for ecotypes and genes in the study. Mean and Standard errors are reported for: N or the number of sequences in the analysis, L or the length removing indels and missing data sites, S or the number of segregating sites, H or haplotype number, Hd or haplotype diversity, π or the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta or the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D or estimated of Tajima's D (* accompany significant values).

Ecotype	loci	L	n		S		H		Hd		π		Theta		D	
			Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Alpine	1174	362	20.000	0.000	7.500	5.500	5.000	2.000	0.442	0.247	0.003	0.002	0.006	0.004	-1.540	0.399
	1176	398	12.000	0.000	1.500	0.500	2.500	0.500	0.243	0.076	0.001	0.000	0.001	0.001	-1.296	0.155
	1192	399	10.000	2.000	1.000	1.000	1.500	1.500	0.357	0.357	0.001	0.001	0.001	0.001	0.414	
	1194	372	22.000	0.000	4.000	1.000	4.000	1.000	0.348	0.171	0.001	0.001	0.003	0.001	-1.619	0.110
	1199	386	16.000	0.000	14.500	11.500	6.000	2.000	0.792	0.075	0.011	0.006	0.014	0.010	-1.043	0.315
	1204	410	12.000	0.000	3.000	2.000	4.000	2.000	0.425	0.258	0.001	0.001	0.003	0.002	-1.334	0.193
	1205	365	14.000	2.000	7.500	3.500	6.000	1.000	0.663	0.087	0.003	0.002	0.006	0.003	-1.732	0.015
	1211	291	20.000	2.000	6.000	2.000	5.500	1.500	0.451	0.118	0.003	0.002	0.006	0.002	-1.414	0.253
	1212	428	22.000	0.000	0.500	0.500	1.000	1.000	0.046	0.046	0.000	0.000	0.000	0.000	-1.162	
Dune	1004	397	11.667	0.803	3.500	0.806	4.000	0.683	0.528	0.123	0.006	0.004	0.003	0.001	-0.984	0.361
	1011	320	20.333	1.202	2.833	1.376	3.333	1.520	0.301	0.160	0.002	0.001	0.002	0.001	-1.160	0.130
	1014	392	21.333	1.116	3.833	0.946	3.833	0.477	0.395	0.088	0.001	0.000	0.003	0.001	-1.296	0.232
	1018	381	17.333	1.687	5.167	2.023	4.167	1.108	0.410	0.111	0.002	0.001	0.004	0.002	-1.464	0.232
	1021	371	18.000	3.307	9.167	3.341	4.667	1.022	0.424	0.104	0.003	0.001	0.007	0.002	-1.847	0.187
	1024	418	22.000	0.730	2.833	1.327	3.167	0.980	0.315	0.075	0.001	0.000	0.002	0.001	-0.800	0.505
	1080	417	21.667	1.202	6.500	3.314	3.833	0.307	0.550	0.061	0.002	0.001	0.004	0.002	-0.819	0.393
	1084	372	13.333	2.044	3.333	1.520	3.500	1.176	0.370	0.096	0.002	0.001	0.003	0.001	-0.997	0.353

Continuation Table S3.3. Polymorphism analysis for ecotypes and genes in the study. Mean and Standard errors are reported for: N or the number of sequences in the analysis, L or the length removing indels and missing data sites, S or the number of segregating sites, H or haplotype number, Hd or haplotype diversity, π or the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta or the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D or estimated of Tajima's D (* accompany significant values).

Ecotype	loci	L	n		S		H		Hd		π		Theta		D	
			Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
	1085	378	17.667	1.202	5.500	1.727	3.000	0.365	0.346	0.099	0.002	0.001	0.004	0.001	-1.629	0.149
	1090	407	18.667	1.116	6.333	1.498	4.000	0.683	0.313	0.079	0.002	0.000	0.005	0.001	-1.749	0.182
	1091	395	11.000	0.683	5.333	1.308	3.833	0.749	0.551	0.096	0.004	0.001	0.005	0.001	-0.468	0.325
	1096	393	12.000	1.461	3.500	1.522	2.500	0.847	0.351	0.122	0.002	0.001	0.003	0.001	-0.981	0.109
	1098	377	22.333	0.615	6.667	1.054	6.500	0.764	0.576	0.084	0.002	0.000	0.005	0.001	-1.603	0.172
	1116	278	19.000	1.125	1.167	0.307	2.000	0.447	0.120	0.032	0.000	0.000	0.001	0.000	-1.302	0.085
	1125	437	12.333	1.202	10.167	2.428	5.500	0.992	0.645	0.094	0.005	0.001	0.008	0.002	-1.601	0.176
	1126	408	13.000	1.528	5.167	2.197	3.500	0.957	0.363	0.132	0.002	0.001	0.004	0.002	-1.389	0.170
	1170	403	11.333	1.033	2.333	1.085	2.500	0.500	0.325	0.106	0.002	0.001	0.002	0.001	-0.637	0.271
	1174	362	17.667	0.803	5.333	0.843	4.667	0.667	0.551	0.068	0.003	0.001	0.004	0.001	-1.026	0.197
	1176	398	12.000	0.000	4.333	1.116	3.833	0.543	0.513	0.077	0.002	0.001	0.003	0.001	-0.626	0.230
Dune	1192	399	15.333	2.667	4.167	1.167	4.000	0.632	0.500	0.088	0.002	0.001	0.003	0.001	-0.784	0.320
	1194	372	22.000	0.730	1.667	0.494	2.333	0.333	0.338	0.071	0.001	0.000	0.001	0.000	0.051	0.586
	1199	386	14.333	0.955	10.167	3.177	6.000	1.390	0.621	0.137	0.009	0.004	0.011	0.004	-1.229	0.392
	1204	410	12.333	0.333	1.333	0.333	2.167	0.477	0.331	0.077	0.001	0.000	0.001	0.000	-0.474	0.406
	1205	365	11.667	1.585	2.167	0.654	2.167	0.477	0.276	0.083	0.001	0.000	0.002	0.000	-1.087	0.177
	1211	291	23.333	1.585	3.333	0.615	4.167	0.601	0.326	0.063	0.001	0.000	0.003	0.001	-1.388	0.204

Continuation Table S3.3. Polymorphism analysis for ecotypes and genes in the study. Mean and Standard errors are reported for: N or the number of sequences in the analysis, L or the length removing indels and missing data sites, S or the number of segregating sites, H or haplotype number, Hd or haplotype diversity, π or the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta or the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D or estimated of Tajima's D (* accompany significant values).

Ecotype	loci	L	n		S		H		Hd		π		Theta		D	
			Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Dune	1212	428	24.000	0.000	2.500	0.500	3.333	0.333	0.223	0.034	0.001	0.000	0.002	0.000	-1.391	0.114
	1004	397	11.667	1.202	4.167	1.537	3.333	0.615	0.482	0.086	0.003	0.001	0.004	0.001	-0.464	0.331
	1011	320	22.000	1.033	2.167	0.601	3.167	0.792	0.308	0.081	0.001	0.000	0.002	0.000	-0.833	0.457
	1014	392	21.667	0.803	3.333	0.715	4.000	0.730	0.311	0.088	0.002	0.000	0.002	0.001	-1.052	0.298
	1018	381	15.667	2.894	6.167	1.621	5.333	1.054	0.569	0.093	0.003	0.001	0.005	0.001	-1.399	0.195
	1021	371	20.000	1.585	11.500	2.742	6.667	0.919	0.711	0.051	0.006	0.002	0.009	0.002	-1.046	0.406
	1024	418	21.667	1.202	3.333	0.919	3.667	0.760	0.363	0.096	0.001	0.000	0.002	0.001	-1.150	0.194
	1080	417	23.000	0.447	2.333	0.919	2.833	0.980	0.385	0.125	0.001	0.000	0.002	0.001	-0.471	0.282
Headland	1084	372	14.667	1.606	3.167	1.046	3.167	0.792	0.408	0.109	0.002	0.001	0.003	0.001	-0.962	0.318
	1085	378	18.333	1.667	2.000	0.365	3.000	0.365	0.405	0.087	0.001	0.000	0.003	0.001	-0.386	0.471
	1090	407	18.333	1.585	6.000	2.463	3.167	0.833	0.306	0.109	0.002	0.001	0.005	0.002	-1.547	0.351
	1091	395	12.000	1.202	6.667	1.406	4.833	0.307	0.632	0.059	0.005	0.002	0.006	0.001	-0.425	0.571
	1096	393	9.667	1.406	3.833	1.249	3.667	0.843	0.662	0.144	0.004	0.001	0.004	0.001	0.120	0.368
	1098	377	19.667	3.556	4.167	1.276	4.000	0.816	0.448	0.135	0.002	0.000	0.003	0.001	-1.689	0.196
	1116	278	22.000	0.803	3.000	1.653	2.000	0.447	0.106	0.028	0.001	0.000	0.003	0.002	-1.720	0.194
	1125	437	11.333	0.803	10.500	3.471	5.000	0.856	0.651	0.064	0.005	0.002	0.008	0.003	-1.127	0.370
	1126	408	12.000	1.633	3.000	1.000	2.833	0.792	0.273	0.074	0.001	0.000	0.002	0.001	-1.610	0.136

Continuation Table S3.3. Polymorphism analysis for ecotypes and genes in the study. Mean and Standard errors are reported for: N or the number of sequences in the analysis, L or the length removing indels and missing data sites, S or the number of segregating sites, H or haplotype number, Hd or haplotype diversity, π or the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta or the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D or estimated of Tajima's D (* accompany significant values).

Ecotype	loci	L	n		S		H		Hd		π		Theta		D	
			Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
	1170	403	12.000	1.789	1.000	0.365	1.667	0.558	0.244	0.100	0.001	0.000	0.001	0.000	-0.562	0.332
	1174	362	19.000	0.683	6.833	1.537	6.000	0.816	0.572	0.089	0.003	0.001	0.005	0.001	-1.364	0.174
	1176	398	12.000	0.000	1.167	0.167	2.167	0.167	0.346	0.040	0.001	0.000	0.001	0.000	-0.124	0.382
	1192	399	17.667	1.713	4.000	0.894	3.667	0.667	0.428	0.094	0.003	0.001	0.003	0.001	-0.330	0.526
	1194	372	23.333	0.422	5.500	2.825	2.500	0.671	0.210	0.070	0.001	0.001	0.004	0.002	-1.609	0.305
	1199	386	14.667	1.116	9.833	2.738	5.500	0.885	0.633	0.092	0.006	0.002	0.009	0.003	-1.513	0.245
Headland	1204	410	13.000	1.983	1.667	0.422	2.167	0.543	0.312	0.072	0.001	0.000	0.001	0.000	-0.478	0.494
	1205	365	12.333	1.202	3.500	0.619	3.500	0.224	0.425	0.050	0.002	0.001	0.003	0.001	-1.481	0.087
	1211	291	23.000	0.447	2.167	1.249	2.000	0.730	0.156	0.074	0.001	0.001	0.002	0.001	-1.196	0.022
	1212	428	24.000	0.000	2.667	0.615	3.333	0.760	0.315	0.092	0.001	0.000	0.002	0.000	-0.981	0.252
	1004	397	7.000	1.000	5.000	2.000	3.000	1.000	0.488	0.155	0.004	0.001	0.005	0.002	-1.296	0.063
	1011	320	21.000	1.000	9.000	3.000	6.500	1.500	0.586	0.114	0.004	0.002	0.008	0.002	-1.564	0.198
	1014	392	21.000	1.000	7.000	3.000	7.000	3.000	0.578	0.215	0.002	0.001	0.005	0.002	-1.709	0.071
Inland	1018	381	11.000	5.000	11.000	4.000	8.500	3.500	0.946	0.013	0.008	0.001	0.010	0.002	-0.475	1.363
	1021	371	21.000	1.000	10.500	1.500	7.000	0.000	0.646	0.125	0.005	0.001	0.008	0.000	-1.541	0.334
	1024	418	22.000	0.000	3.000	1.000	3.500	0.500	0.255	0.078	0.001	0.000	0.002	0.001	-1.591	0.076
	1080	417	22.000	0.000	11.500	4.500	7.500	2.500	0.814	0.065	0.005	0.001	0.008	0.003	-1.270	0.638

Continuation Table S3.3. Polymorphism analysis for ecotypes and genes in the study. Mean and Standard errors are reported for: N or the number of sequences in the analysis, L or the length removing indels and missing data sites, S or the number of segregating sites, H or haplotype number, Hd or haplotype diversity, π or the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta or the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D or estimated of Tajima's D (* accompany significant values).

Ecotype	loci	L	n		S		H		Hd		π		Theta		D	
			Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
	1084	372	14.000	2.000	4.000	1.000	5.000	1.000	0.600	0.158	0.002	0.001	0.003	0.001	-1.438	0.089
	1085	378	16.000	5.000	10.000	8.000	4.500	1.500	0.410	0.194	0.004	0.003	0.008	0.006	-1.796	0.289
	1090	407	8.000	2.000	4.500	3.500	4.000	2.000	0.733	0.197	0.004	0.002	0.004	0.003	-0.127	1.294
	1091	395	11.000	3.000	7.000	3.000	5.500	2.500	0.765	0.158	0.006	0.001	0.007	0.002	0.006	0.580
	1096	393	9.000	3.000	6.500	3.500	4.500	2.500	0.606	0.273	0.006	0.003	0.007	0.004	-0.989	0.244
	1098	377	22.000	0.000	10.500	2.500	9.000	2.000	0.650	0.109	0.004	0.000	0.009	0.000	-2.020	0.194
Inland	1116	278	21.000	1.000	1.500	1.500	2.000	2.000	0.142	0.142	0.001	0.001	0.002	0.002	-1.723	
	1125	437	10.000	0.000	5.000	0.000	5.500	0.500	0.823	0.045	0.004	0.001	0.004	0.000	-0.556	0.227
	1126	408	11.000	1.000	2.500	1.500	3.500	1.500	0.466	0.110	0.002	0.001	0.002	0.001	-0.685	0.700
	1170	403	10.000	0.000	6.500	0.500	5.000	0.000	0.667	0.000	0.004	0.000	0.006	0.000	-1.858	0.019
	1174	362	14.000	3.000	4.500	1.500	4.000	0.000	0.606	0.045	0.003	0.001	0.004	0.002	-1.023	0.609
	1176	398	10.000	0.000	2.500	1.500	3.000	1.000	0.367	0.167	0.001	0.001	0.002	0.001	-1.452	0.340
	1192	399	18.000	2.000	5.500	3.500	5.000	2.000	0.513	0.271	0.003	0.002	0.004	0.003	-1.295	0.203
	1194	372	22.000	7.000	6.500	3.500	4.000	0.000	0.260	0.000	0.002	0.001	0.005	0.003	-1.954	0.224
	1199	386	15.000	1.000	18.000	9.000	7.500	1.500	0.831	0.062	0.013	0.004	0.017	0.006	-0.937	0.356
	1204	410	8.000	0.000	2.500	0.500	4.500	1.500	0.750	0.143	0.004	0.002	0.005	0.003	-1.086	0.638
	1205	365	14.000	0.000	4.500	1.500	5.000	1.000	0.550	0.055	0.002	0.000	0.004	0.001	-1.262	0.697

Table S3.3. Polymorphism analysis for ecotypes and genes in the study. Mean and Standard errors are reported for: N or the number of sequences in the analysis, L or the length removing indels and missing data sites, S or the number of segregating sites, H or haplotype number, Hd or haplotype diversity, π or the estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987), Theta or the Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975) and D or estimated of Tajima's D (* accompany significant values).

Ecotype	loci	L	n		S		H		Hd		π		Theta		D	
			Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Inland	1211	291	22.000	0.000	3.000	1.000	4.500	1.500	0.294	0.117	0.001	0.001	0.003	0.001	-1.756	0.241
	1212	428	22.000	0.000	0.500	0.500	1.000	1.000	0.087	0.087	0.000	0.000	0.000	0.000	-0.641	

Table S3.4. AMOVA and Hierarchical analyses for each of the neutral genes in the study in *Senecio lautus.* Variances (Var) are estimated Among Groups (AG), Among Populations (AP) and Within Populations (WP). Fixation indices (F_{SC}, F_{ST} and F_{CT}) are reported with the corresponding significance. Negative variance values and fixation indices indicte that there is no genetic structure for locus 1176.

Locus	Var AG	Var AP	Var WP	F_{SC}	P-F_{SC}	F_{ST}	P-F_{ST}	F_{CT}	P-F_{CT}
1011	1.147	7.217	91.636	0.073	0.000	0.084	0.000	0.011	0.075
1014	4.735	6.891	88.374	0.072	0.000	0.116	0.000	0.047	0.002
1024	1.651	3.495	94.854	0.036	0.000	0.051	0.000	0.017	0.006
1080	9.189	11.579	79.232	0.128	0.000	0.208	0.000	0.092	0.000
1096	-0.481	3.015	97.466	0.030	0.000	0.025	0.000	-0.005	0.719
1116	-0.493	3.508	96.985	0.035	0.006	0.030	0.008	-0.005	0.588
1174	0.917	1.291	97.792	0.013	0.005	0.022	0.000	0.009	0.013
1176	-0.157	-1.148	101.305	-0.011	0.967	-0.013	0.992	-0.002	0.845
1192	1.563	9.535	88.901	0.097	0.000	0.111	0.000	0.016	0.106
1194	2.266	8.192	89.542	0.084	0.000	0.105	0.000	0.023	0.083
1199	0.338	1.486	98.176	0.015	0.006	0.018	0.002	0.003	0.161

Continuation Table S3.4. AMOVA and Hierarchical analyses for each of the neutral genes in the study in *Senecio lautus*. Variances (Var) are estimated Among Groups (AG), Among Populations (AP) and Within Populations (WP). Fixation indices (F_{SC}, F_{ST} and F_{CT}) are reported with the corresponding significance. Negative variance values and fixation indices indicte that there is no genetic structure for locus 1176.

Locus	Var AG	Var AP	Var WP	F_{SC}	P-F_{SC}	F_{ST}	P-F_{ST}	F_{CT}	P-F_{CT}
1211	6.379	10.100	83.521	0.108	0.000	0.165	0.000	0.064	0.002
1212	19.478	15.583	64.939	0.194	0.000	0.351	0.000	0.195	0.000

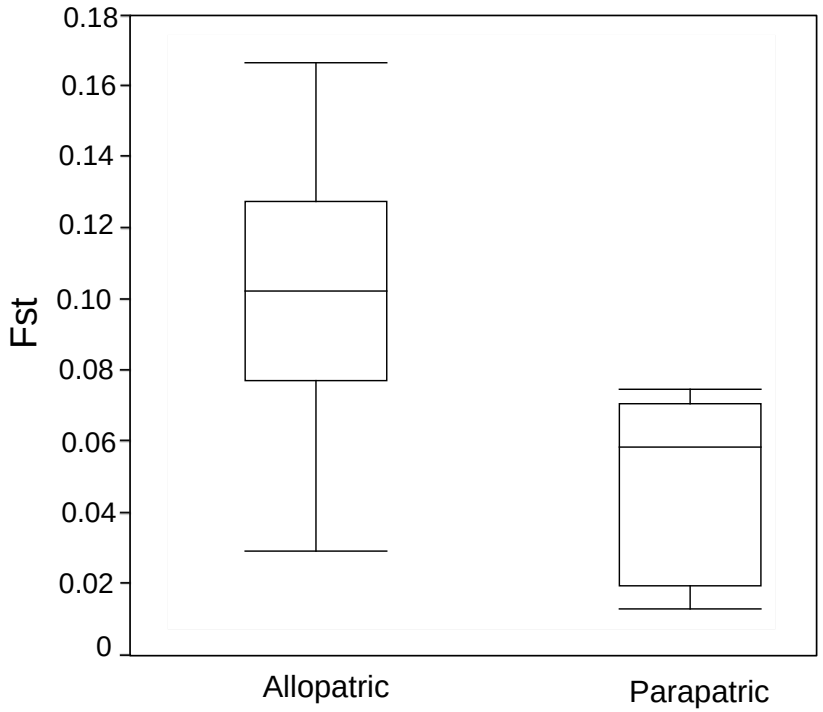

Fig. S3.1. Fst values for allopatric and parapatric comparisons of Dune and Headland populations in *Senecio lautus.*

SUPPLEMENTARY CHAPTER IV

Table S4.1. Populations of *Senecio lautus* crossed and number of families produced for each cross type.

	D01	D04	D23	D32	H01	H05	H12	H21
D01	12	14	16	4	10		2	
D04		13	16	6	2	10	1	
D23			13	6	7	1		17
D32				5			9	
H01					19		4	12
H05						16	10	
H12							6	11
H21								12

Table S4.2. The Strength of RI for multiple population comparisons of *Senecio lautus* and three RI barriers: Immigrant inviability (RI 1), F1 seed set (RI 2) and F1 inviability (RI 3).

Region	Ecotype	Comparison	RI 1	RI 2	RI 3
Within	Different ecotype	D01-H05	0	-	-
Between	Different ecotype	D01-H21	-0.153846154	-	-
Between	Different ecotype	D32-H01	0.060150376	-	-
Between	Different ecotype	D32-H05	0.107142857	-	-
Within	Different ecotype	D32-H21	0.175824176	-	-
Between	Different ecotype	H01-D32	-0.773333333	-	-
Within	Different ecotype	H05-D01	-0.533333333	-	-
Between	Different ecotype	H05-D32	-0.866666667	-	-
Between	Same ecotype	H05-H21	-0.461538462	-	-
Within	Different ecotype	H12-D23	-0.062091503	-	-
Between	Different ecotype	H21-D01	0.133333333	-	-
Between	Different ecotype	H21-D04	0.15	-	-

Continuation Table S4.2. The Strength of RI for multiple population comparisons of *Senecio lautus* and three RI barriers: Immigrant inviability (RI 1), F1 seed set (RI 2) and F1 inviability (RI 3).

Region	Ecotype	Comparison	RI 1	RI 2	RI 3
Within	Different ecotype	H21-D32	-0.306666667	-	-
Between	Same ecotype	H21-H05	0	-	-
Between	Same ecotype	D04-D32	-0.773333333	-0.041254419	-
Between	Different ecotype	D04-H21	-0.307692308	1	-
Within	Same ecotype	D23-D32	-0.4	-0.155018099	-
Within	Different ecotype	D23-H12	0.018181818	0.843265937	-
Between	Same ecotype	D32-D04	0.285714286	-0.329864823	-
Within	Same ecotype	D32-D23	0.159663866	-0.063659212	-
Between	Different ecotype	H05-D23	-0.352941176	1	-
Between	Different ecotype	H12-D04	0.027777778	1	-
Within	Same ecotype	D01-D04	0.15	-0.426256353	-0.134674395
Between	Same ecotype	D01-D23	-0.294117647	-0.066631428	-0.578947368

Continuation Table S4.2. The Strength of RI for multiple population comparisons of *Senecio lautus* and three RI barriers: Immigrant inviability (RI 1), F1 seed set (RI 2) and F1 inviability (RI 3).

Region	Ecotype	Comparison	RI 1	RI 2	RI 3
Between	Same ecotype	D01-D32	-0.586666667	-1.386370855	-0.059602649
Within	Different ecotype	D01-H01	0.210526316	-0.269729378	-0.036899225
Between	Different ecotype	D01-H12	0.410909091	0.583761864	0.280898876
Within	Same ecotype	D04-D01	-0.266666667	-0.290385523	0.007159905
Between	Same ecotype	D04-D23	-0.176470588	-0.102460981	-0.460144928
Within	Different ecotype	D04-H01	0.263157895	0.824006968	-0.105640107
Within	Different ecotype	D04-H05	-1.125	-0.698700507	0.072611465
Between	Different ecotype	D04-H12	0.083636364	0.96630114	1
Between	Same ecotype	D23-D01	0.4	0.225014935	0.210526316
Between	Same ecotype	D23-D04	0.5	-0.188928563	0.057971014
Between	Different ecotype	D23-H01	0.473684211	0.948167137	0.418960245
Between	Different ecotype	D23-H05	0.125	0.86762896	0.565217391

Continuation Table S4.2. The Strength of RI for multiple population comparisons of *Senecio lautus* and three RI barriers: Immigrant inviability (RI 1), F1 seed set (RI 2) and F1 inviability (RI 3).

Region	Ecotype	Comparison	RI 1	RI 2	RI 3
Within	Different ecotype	D23-H21	-0.153846154	0.001456752	0.422740525
Between	Same ecotype	D32-D01	0.285714286	0.233258628	-0.324503311
Within	Different ecotype	D32-H12	0.298701299	-0.836578044	-0.000744602
Within	Different ecotype	H01-D01	-0.266666667	-0.331723672	0.057364341
Within	Different ecotype	H01-D04	0.05	0.664420066	-0.105640107
Between	Different ecotype	H01-D23	0.058823529	0.889910176	0.47706422
Within	Same ecotype	H01-H05	-1	-0.019566385	0.006142263
Between	Same ecotype	H01-H12	0.018181818	0.32148155	-0.071743346
Between	Same ecotype	H01-H21	-0.153846154	0.628820063	-0.022994952
Within	Different ecotype	H05-D04	-0.05	-0.409450211	0.072611465
Within	Same ecotype	H05-H01	0.052631579	-0.116382809	0.781652467
Between	Same ecotype	H05-H12	0.083636364	0.550598332	0.143812709

Continuation Table S4.2. The Strength of RI for multiple population comparisons of *Senecio lautus* and three RI barriers: Immigrant inviability (RI 1), F1 seed set (RI 2) and F1 inviability (RI 3).

Region	Ecotype	Comparison	RI 1	RI 2	RI 3
Between	Different ecotype	H12-D01	-0.111111111	0.899479906	0.400749064
Within	Different ecotype	H12-D32	-0.425925926	-0.966712011	-0.012658228
Between	Same ecotype	H12-H01	0.049707602	0.612564518	0.249779658
Between	Same ecotype	H12-H05	-1.083333333	0.472489845	0.143812709
Within	Same ecotype	H12-H21	-0.709401709	-0.466798079	-0.259158752
Within	Different ecotype	H21-D23	-0.586666667	0.514029283	-0.142857143
Between	Same ecotype	H21-H01	0.315789474	0.527617373	0.392596747
Within	Same ecotype	H21-H12	0.214545455	-1.337641513	0.218453189